7-16-2014

An Integrated Approach to Understanding the Structure and Function of the Large-Conductance, Voltage- and Calcium-Activated Potassium Channel

Jacob E. McMillan

AN INTEGRATED APPROACH TO UNDERSTANDING THE STRUCTURE AND FUNCTION OF THE LARGE-CONDUCTANCE, VOLTAGE- AND CALCIUM-ACTIVATED POTASSIUM CHANNEL

by

Jacob Evan McMillan

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

Major: Chemistry

The University of Memphis

August 2014

Dedication

It is in the moments we are celebrating our accomplishments that we must recognize they are only possible because of those who came before us. We can transcend the accomplishments of our predecessors only because of what they have given us through the propagation of knowledge. To the many teachers and mentors throughout my education, my family and friends, and most importantly, my parents, Harold and Kathryn McMillan, who have shown unwavering support of my academic endeavors: I owe all of you and the generations of scientists who came before me a sincere debt of gratitude. I dedicate this work to all of you because without you it would not be attainable.

nanos gigantum humeris insidentes

Acknowledgements

Portions of this work were supported by, and would not be possible without, awards from the National Institutes of Health (Award Numbers R01-HL104631 and R37-AA11560). A generous donation of licenses by The Chemical Computing Group made work with the Molecular Operating Environment possible. I want to thank Dr. Abby L. Parrill, Dr. Daniel L. Baker, and Dr. Alexandra M. Kikonyogo for their years of mentorship, advice, and patience. I would also like to thank all of my research group members, both past and present, for assistance throughout the years. Finally, I would like to thank Dr. Alex M. Dopico and Dr. Anna N. Bukiya at the University of Tennessee Health Science Center for their mentorship and for making this tremendously rewarding collaboration possible.

Abstract

McMillan, Jacob Evan. PhD. The University of Memphis. August 2014. An Integrated Approach to Understanding the Structure and Function of the Large-Conductance, Voltage- and Calcium-Activated Potassium Channel. Major Professor: Abby L. Parrill.

The large-conductance, voltage- and calcium-activated potassium channel (BK_{Ca} channel) is an important transmembrane ion channel involved in many physiological and pathophysiological processes. However, little is known about its atomic-level three-dimensional structure. The functional channel requires four alpha subunits assembled as a potassium pore and is typically associated with four beta subunits that regulate function in various ways depending on the isoform of the beta subunit is interacting with the channel. Currently, the only major high-resolution atomic structure available is of the cytosolic tail domain of the alpha subunit. In the absence of structural information, electrophysiology has allowed for extensive characterization of the channel's role in physiological processes and advanced understanding of BK_{Ca} channel pharmacology including inhibition of BK_{Ca} by physiologically relevant concentrations of cholesterol and stimulation of BK_{Ca} by the related sterol, lithocholic acid. However, many questions still remain including how does cholesterol exert effects on the alpha subunit, what are the structures of the different beta subunits, and can therapeutic lead compounds be developed to selectively target BK_{Ca} beta subunits in a tissue specific manner. The difficulty of obtaining transmembrane protein structures has made characterization a major challenge.

Herein, work is described where 1) the BK_{Ca} beta 1 subunit was successfully expressed and purified from *E. coli* for the first time with a yield of

approximately 32mg/L of bacterial growth and several assays were attempted to demonstrate protein functionality. However, functional assessment of purified protein has remained elusive due to the hydrophobic nature of known ligands. Additional effort is needed to establish ligand recognition using solution nuclear magnetic resonance spectroscopy. 2) Multiple generations of pharmacophore models have been developed to aid in virtual screening efforts to find therapeutic lead compounds targeting BK_{Ca} through the beta 1 subunit. Sixteen compounds have been selected from virtual screening of the PubChem database that can be tested *in vitro* for model validation. 3) Molecular dynamics simulations were utilized to study the underlying mechanism of the interaction between cholesterol and the BK_{Ca} alpha subunit cytosolic tail domain showing differential behavior of cholesterol in mutant and wild type simulations that agree with experimental data.

Table of Contents

Appendices

List of Tables

List of Figures

List of Abbreviations and Acronyms

3-HENA	3-hydroxyolean-12-en-30-oic acid
3D	Three dimensional
6x-His tag	Hexahistidine fusion tag
BK_{Ca}	Large-conductance, voltage- and calcium-gated potassium channel
BMS	Bristol Meyers Squibb
bp	Base pair
BSA	Bovine serum albumin
BSA-E	β-estradiol 6-(O-carboxy-methyl)oxime: BSA
BSA-E-F	β-estradiol 6-(O-carboxy-methyl)oxime: BSA fluorescein iso-thiocyanate
CADD	Computer aided drug design
CLR	Cholesterol
CRAC	Cholesterol recognition amino acid consensus
CTD	Cytosolic tail domain
D	Aspartate
DHPC	1,2-diheptanoyl-*sn*-glycero-3-phosphocholine
DHS-1	Dehydrosoyasaponin-1
DNA	Deoxyribonucleic acid
DPC	n-dodecylphosphocholine
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
F	Phenylalanine
HSQC	Heteronuclear single quantum coherence
IMAC	Immobilized metal affinity chromatography
IK_{Ca}	Intermediate-conductance calcium-gated potassium channel

IPTG	Isopropyl β-D-1-thiogalactopyranoside
K	Lysine
K_d	Dissociation rate constant
kDa	Kilodaltons
L	Leucine
LB	Lysogeny broth
LCA	Lithocholic acid
LIC	Ligation independent cloning
LJ	Lennard-Jones
MBP	Maltose binding protein
MD	Molecular dynamics
MOE	Molecular Operating Environment
MRC	Molecular Resource Center
mRNA	Messenger ribonucleic acid
NCBI	National Center for Bioinformatics
NCI	National Cancer Institute
NEB	New England Biolabs, Inc.
NMR	Nuclear magnetic resonance
OD	Optical density
Oligo	Oligonucleotide
PBS	Phosphate buffered saline
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PME	Particle Mesh Ewald
pS	PicoSiemens

R	Arginine
RMSD	Root mean squared deviation
RNA	Ribonucleic acid
rpm	Revolutions per minute
SAR	Structure activity relationship
Sarcosyl	N-lauroylsarcosine
SASA	Solvent accessible surface area
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SFU	Standardized fluorescence units
SK_{Ca}	Small-conductance calcium-gated potassium channel
SOC	Super optimal broth with catabolite repression
SPR	Surface plasmon resonance
SUMO	Small ubiquitin-like modifier
T	Threonine
TAE	Tris base, acetic acid, and EDTA
TEV	Tobacco etch virus
TM	Transmembrane
TVMV	Tobacco vein mottling virus
UTHSC	University of Tennessee Health Science Center
V	Valine
WT	Wild type
Y	Tyrosine

Chapter 1

An Introduction to Experimental Techniques and the Large Conductance, Voltage- and Ca^{2+}-Activated K^+ Channel

The large conductance, voltage- and Ca^{2+}-activated K^+ channel, also called BK, BK_{Ca}, MaxiK or *Slo1* channel, is a voltage- and Ca^{2+}-gated K^+ ion channel that is ubiquitously expressed and localized in mammalian cell membranes.[1] It is a member of the six transmembrane domain (TM6) ion channel super family and the Ca^{2+}-activated K^+ channel family, also called the K_{Ca} family. Although all members of the K_{Ca} family (BK_{Ca}, IK_{Ca} and SK_{Ca}) are involved in repolarization of cell membrane potential, BK_{Ca} plays a role in several important physiological events including neuronal hyperpolarization, neurotransmission, hormonal secretion, and regulation of tone in both vascular and non-vascular smooth muscle.[2,3] All members of the K_{Ca} family share structural features including their voltage-sensing domains, which open the K^+ pore in response to depolarization of the cell membrane, and pore-forming transmembrane domains (to be discussed later). Yet they vary in, and are classified by, their conductance. Of all known Ca^{2+}-gated K^+ channels family members, the BK_{Ca} channel, which stands for big conductance of K^+, has the largest unitary conductance at 100-300 picoSiemens (pS)[4] while other members, including the intermediate conductance (IK_{Ca}) and small conductance (SK_{Ca}) channels, have conductances of 25-100 pS[5,6] and 2-25 pS[7] respectively. While the effect of Ca^{2+} concentration on cellular K^+ permeability was first reported in 1958[5] and a change in ionic current due to increased intracellular calcium was

demonstrated in 1970,[8] it was not until 1981[4] that the BK_{Ca} channel was identified as the first member of the K_{Ca} family partially responsible for the observed effect that increased intracellular Ca^{2+} has on K^+ permeability.

1.1 Structure and Function of the α and β Subunits

Functional BK_{Ca} channels result from association of four identical slo1 proteins, also called α subunits[9], which were first characterized from *Drosophila* by identifying conserved sequences between BK_{Ca} and the previously characterized *Shaker* voltage-gated potassium channel.[10,11] Each α subunit is composed of seven transmembrane (TM) segments (Figure 1, labeled S0 through S6), which form both the voltage-sensing domain and the pore-forming domain. The voltage-sensing machinery includes elements that span the S0 to S4 segments while the pore-forming domain is comprised of S5 and S6[12] (Figure 1). Voltage sensing is primarily conferred by a series of charged residues on different segments in the voltage-sensing domain that will alter protein conformation in response to ion concentration changes in the intracellular or extracellular environment through charge repulsion; D153 and R167 on S2, D186 on S3, and R213 on S4.[13] The S0 segment and the long cytosolic tail domain (CTD) are found in BK_{Ca}, but not in purely voltage-gated TM6 potassium channels and are required for modulation by accessory proteins termed BK_{Ca} β subunits.[14-16]

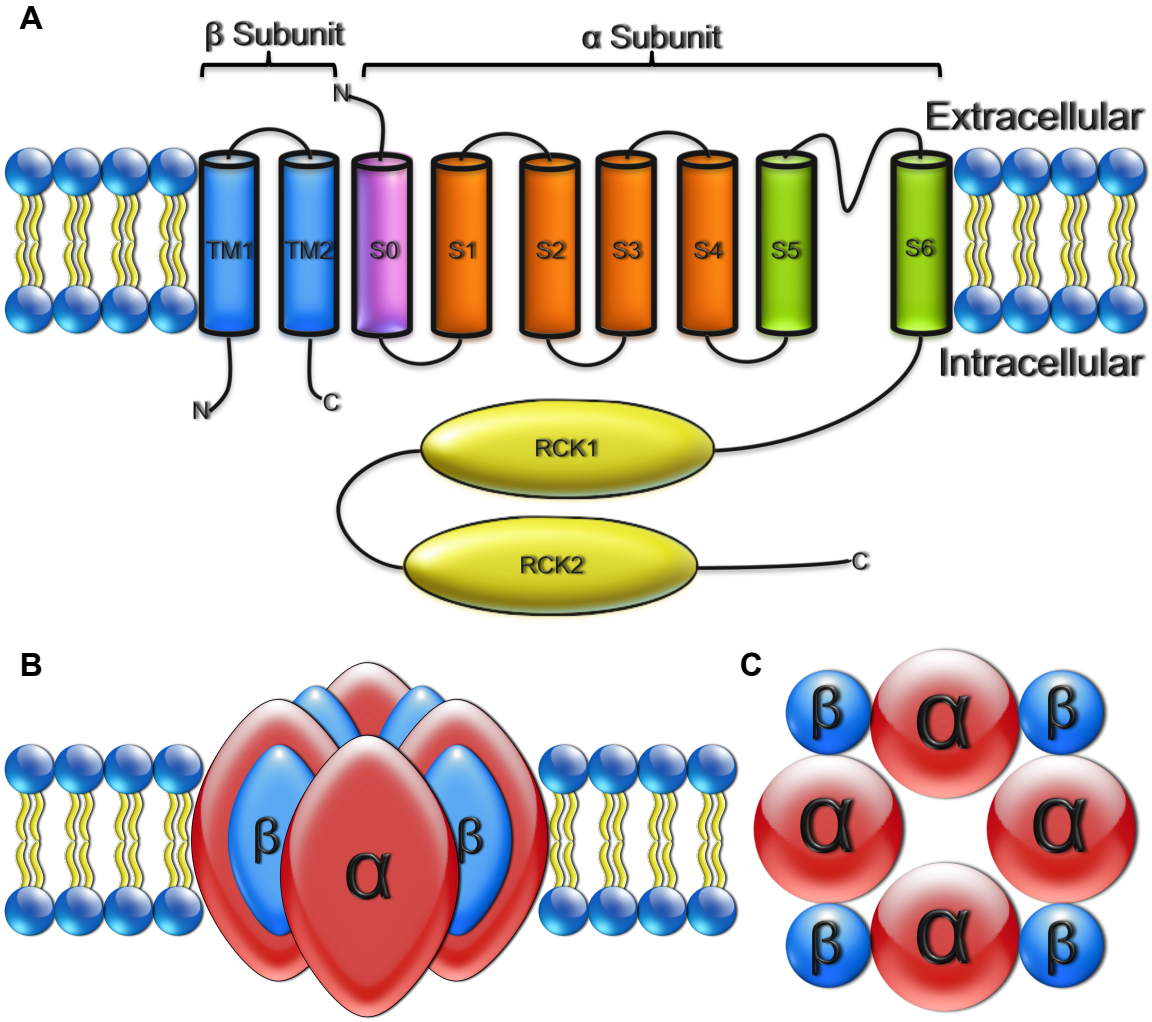


Figure 1. Cartoon Diagram of BK_{Ca} α and β Subunits
A) The pore-forming α subunit contains seven TM domains (S0-S6) and a long intracellular tail domain (CTD) where two Regulator of Conductance for K^+ (RCK) domains are found. S0 (purple) is a primary element for interaction between α and β subunits. Structural elements from S0 to S4 (orange) participate in the voltage-sensing machinery while S5 and S6 (green) form the K^+ pore. The β subunit (blue) consists of two TM domains with an extracellular loop connecting the TM domains and has short intracellular C- and N-termini. B) The functional channel is composed of four identical α subunits (colored red), which can operate alone or with four regulatory β subunits (colored blue). C) Top view of BK_{Ca} channel assembly. β subunits (blue) interface between two pore-forming α (red) subunits to complete the hetero-octomeric quaternary structure of the channel.

The long CTD contains approximately 800 residues and includes two regulator of conductance for K^+ domains (RCK1 and RCK2).[17] These domains include two high-affinity Ca^{2+} binding sites that embolden the channel to open at a lower membrane potential threshold in response to increase in Ca^{2+} within the intracellular physiological range (hundreds of nanomolar to tens of micromolar) in a concentration dependent manner (as Ca^{2+} concentrations increase, the voltage required to open the channel decreases).[18] The proximal high affinity site is located in RCK1 and involves residues D362 and D367[19] and the distal is in RCK2 and includes the so-called "Ca^{2+}-bowl", a penta-aspartate sequence.[20]

In smooth muscle, BK_{Ca} channels induce currents that result from channel activation following the release of Ca^{2+} from intracellular Ca^{2+} stores or influx of Ca^{2+} through Ca_v channels after initial membrane depolarization. Intracellular Ca^{2+} in turn binds to and opens BK_{Ca} channels leading to hyperpolarization, or a negative electrical potential shift, of the cell membrane by allowing K^+ to flow down its electrochemical gradient and out of the cell.[21-24] Hyperpolarization of the membrane then prevent further Ca^{2+} influx into the cell by deactivating voltage-dependent Ca^{2+} channels, which remain closed at the negative resting potential of approximately -70 mV, effectively terminating the influx of Ca^{2+} and stopping muscle contraction.[25,26]

While differential phenotypical channel gating and kinetics can be achieved through alternative splicing of the *slo1* gene to produce different isoforms of the α subunit,[27-29] accessory β subunits offer alternative means of channel modulation. Expression of the four known accessory β subunits (β_1-β_4) is

highly tissue-specific, and their presence in the BK_{Ca} channel complex provides a tissue-specific phenotype (see below) that allows the BK_{Ca} channel to serve cell physiology in a tissue-specific manner.[30-32] The β subunits are smaller proteins composed of two putative TM segments, intracellular N- and C-termini, and an extracellular loop connecting TM1 and TM2 that is 116-128 amino acids in length[30,33] (Figure 1). Four β subunits assemble with the tetrameric α channel to form a hetero-octameric quaternary structure.[34] Crosslinking experiments indicate that TM1 of the β subunit is positioned near S1 of one α subunit and TM2 is positioned near S0 in another α subunit of the assembled channel.[35] A recent mutagenesis study suggests that β_1 and β_2 subunits may increase channel opening at lower voltages as $[Ca^{2+}]$ increases by disrupting the disulfide bonds formed at interface between the voltage-sensing domain and the intracellular C-terminal domain and also decrease voltage sensitivity by preventing ionic interactions with R213 in the voltage-sensing domain.[36]

Each subunit has distinctly different effects on channel behavior. β_1 has the largest effect on voltage-dependent activation, both reducing the voltage sensitivity by shifting the voltage required to open 50% of channels ($V_{1/2}$) in patch clamp experiments more negative[37,38] and increasing Ca^{2+} sensitivity by reducing the voltage required to open the channel as $[Ca^{2+}]$ was increased in patch clamp experiments.[39] β_1 is expressed primarily in smooth muscle and the kidneys.[40-42] The β_2 subunit does not shift on voltage-dependent opening and closing of the channel,[37,38] but makes the channel more sensitive to Ca^{2+} concentration as evidenced by the decreased voltage required to open the channel as $[Ca^{2+}]$

increases,[43] although not to the same extent as β_1, and can decrease channel open probability channel conductance through an intracellular N-terminal “ball” peptide.[44] β_2 is expressed primarily in the brain.[44] The two subunits appear to confer Ca^{2+} sensitivity in different ways as mutations in the voltage-sensing domain that reduce the voltage sensitivity modulation by β_1 had no effect on modulation by β_2.[45] The β_3 subunit, which is the least well studied, also inactivates the channel through an N-terminal peptide[46], but not to the same extent as β_2.[47] This “incomplete” inactivation could be due to the lower affinity of BK_{Ca} α for β_3 when compared to β_2, as evidenced by faster dissociation rates for β_3.[48] β_3 is expressed in many tissues including the testis[49] and the pancreas.[50] The β_4 subunit has been shown to slow gating kinetics while also increasing Ca^{2+} sensitivity,[31] and is expressed in neuronal tissues.[49] Finally, several new modulatory proteins, called leucine-rich repeat-containing protein, or “BK_{Ca} γ subunits,” have been shown shift the voltage dependence of the channel to -140 mV, representing the largest known modulation of the voltage dependence of BK_{Ca} α.[51,52] However, little else is known about these accessory proteins at this time.

While no complete crystal structure of the α subunit has been published to date, the structure of the intracellular gating ring (shown as RCK1 and RCK2 in Figure 1 and also called the cytosolic tail domain or CTD) of the BK_{Ca} channel α subunit has been determined in both the open and closed “states” by X-ray crystallography.[17,53] The transmembrane segments, however, remain largely uncharacterized. Recently, the intracellular loop between S0 and S1 was

characterized by NMR and shown to contain two amphipathic helices, which are expected considering their proximity to the cell membrane.[54] In addition, a cryo-electron microscopy structure of the whole channel has been reported showing a globular structure similar, in many ways, to K^+ channels of known structure.[55]

Even less high-resolution structural data are available for the β subunits: only the structure of the N-terminal peptide of β_2 determined by NMR has been reported to date.[56] The lack of atomic resolution structural data leaves a tremendous gap in our understanding of both the structure and function of a channel that is involved in critical aspects of mammalian physiology, as mentioned in section 1.3. Thus, further structural studies would greatly enhance our understanding of BK_{Ca} channel activation and deactivation through examination of structures in both states, modulation by Ca^{2+} and voltage by identifying the structural features and changes involved, interactions with different β subunits, and identification of binding and interaction sites with ligands. Expanding understanding of the topics could lead to the development of novel therapeutic lead compounds. For a recent and extensive review on BK_{Ca} structure, function, and physiology, which will be discussed to an extent in the coming sections, see Contreras, 2013[57] or Dopico *et al.* 2012.[58]

1.2 The β_1 Subunit

The β_1 subunit, which is the focus of Chapter 3 and Chapter 4 of this dissertation, is the best studied of the β subunits. When the α subunits are expressed alone, the channel has low sensitivity to Ca^{2+} and shows activation only at highly positive membrane potentials of +40 mV[59] like those seen after membrane depolarization as a result of action potentials in neurons and muscles.

In the presence of the β_1 subunit, however, the channel exhibits increased Ca^{2+} and decreased voltage sensitivity.[60,61] In addition, both channel opening and closing kinetics are slowed in patch clamp recordings.[62] Modulation of voltage sensitivity and gating of BK_{Ca} by β_1 is known to involve four residues in the extracellular loop of β_1 (Y74, S104, Y105, I106), which have been identified by alanine-scanning mutagenesis[63] in addition to the possible disruption of α subunit domains previously mentioned.[36] The CTD was also identified as essential for β_1 modulation of the channel as CTD truncated channels are not modulated by β_1.[64] However, the interaction and interfacing between α and β subunits is not understood.

The β_1 subunit has a well-characterized cholane steroid-recognition site,[65] which will be discussed in more detail in section 1.4 and is not found in any other β subunits, providing this subunit with distinct pharmacological properties. Moreover, the tissue-specific expression of β_1 in smooth muscle[40] and the role β_1 plays in vasoregulation by inducing vasodilation[41] makes it an attractive target for structural characterization. This would provide understanding of the binding and interaction sites for in the β_1 subunit and could lead to development of agents to counteract pathophysiologies, discussed in the next section, by targeting β_1-containing BK_{Ca} channels (e.g., new vasodilators or smooth muscle relaxants).

1.3 Pathophysiological Relevance of the Large-Conductance Ca^{2+}-Activated K^+ Channel

The BK_{Ca} channel plays a role in membrane polarization events through Ca^{2+} and voltage gating, or the opening and closing of the K^+ pore,[66] which can act independently. As a result, this channel is a critical player in many

physiological processes including insulin secretion,[67] smooth muscle dilation,[30] neuronal transmission,[68,69] arterial constriction and dilation,[25,70] cochlear hair cell tuning,[71] detection of O_2 concentration in the carotid body,[72] release of glutamate from nerve terminals,[73] capacitation of human spermatozoa,[74] apoptosis[75] and cell viability,[76] K^+ secretion from the kidneys,[77] and signaling by luteinizing hormone and human chorionic gonadotropin in testicular Leydig cells.[78] As mentioned previously, tissue specific channel behavior is controlled through slo1 gene splicing and differential, tissue-specific expression of β subunit types allowing the channel to be involved in so many different physiological processes.

Mutations and alterations in BK_{Ca} expression can result in pathological conditions. In particular, several neurological conditions appear to involve the BK_{Ca} channel. For example, one study showed that mRNA levels of BK_{Ca} *slo1* transcripts are up to 40% lower in the prefrontal cortex of people diagnosed with schizophrenia.[79] Another study showed that haploinsufficiency (having only one copy of a gene) and a mutation in a highly conserved region of the α subunit of BK_{Ca} cause the decreased channel activity found in people with autism and mental retardation.[80] It was postulated that the mutation and decrease in channel activity, which reduce neuronal excitability and synaptic transmissions, contributed to the pathophysiology of these disorders.[80] Furthermore, deletion of A750 in the β_3 subunit gene, which causes early termination of the protein during translation, is correlated to people with idiopathic generalized epilepsy.[81] The D434G mutation in the α subunit has also been linked to epilepsy and a movement disorder.[82] This may be caused by the increased Ca^{2+} sensitivity and

open time characteristic of the mutant channel,[82-85] which are due to increased flexibility in the intracellular portions of the protein.[86,87] Additionally, BK_{Ca} β_4 knockout mice show epileptic repetitive action potentials in electroencephalograms, but do not exhibit the entire epileptic phenotype.[88] It is hypothesized that β_4 serves to reduce the contribution of the BK_{Ca} channels to membrane repolarization.[88]

Several pathologies related to smooth muscle function are also linked to the BK_{Ca} channel. Genetic knockout[41] or suppression[89] of the β_1 gene in mice produces animals with systemic hypertension. This is caused by the loss of the modulatory effect of β_1 on the channel-forming α proteins; this loss makes the channel complex insensitive to ryanodine receptor-mediated Ca^{2+} "sparks" and prevents vasodilation,[90,91] but may also cause the loss of ability to secrete K^+ from the kidneys.[92] Conversely, a gain-of-function mutation in human β_1 showed protection against diastolic hypertension in a population epidemiology study.[93] For a review on the role BK_{Ca} plays in the regulation of vascular tone, see Ledoux 2006.[94] Erectile dysfunction[95] and hypercontractility of corpus cavernosum smooth muscle[96] have also been reported in *Slo1* genetic knockout mice. *Slo1* knockout mice also displayed overactive bladder and incontinence caused by increased bladder smooth muscle contractility from the lack of current produced by BK_{Ca}.[97,98] Furthermore, a recent study evaluated expression levels of α and β subunits in human urinary bladder and found decreased expression of both subunits in smooth muscle of people with overactive bladder.[99] In the lungs, the role of BK_{Ca} in bronchial smooth muscle is less well understood. However, β_1

subunit knockout mice showed increased tracheal contractility compared to controls,[100] as normal BK_{Ca} works by opposing the effects of cholinergic M2 receptor activity in tracheal smooth muscle cells.[101] Additionally, activated BK_{Ca} channels in lung endothelial cells have been shown to induce vasodilation.[102]

Finally, BK_{Ca} α and β subunits play roles in kidney function, diabetes, and obesity. The α and all four β subunits are expressed in the nephrons[103] with α subunit expression being regulated by aldosterone levels and urine pH.[104] In the kidneys, BK_{Ca} serves to secrete K^+ from the distal nephron.[77] Diets high in K^+ have been shown to increase expression of BK_{Ca} α in the kidneys, while low K^+ diets decreased expression in immunohistochemical staining[105] and, as mentioned previously, the systemic hypertension reported in β_1 knockout mice might be a result of increased aldosterone levels and decreased K^+ secretion efficiency.[92] Increases in glomerular filtration rate and Na^+ secretion, as well as decreases in K^+ secretion, were also observed in β_1 knockout mice.[42] For an extensive review of the role BK_{Ca} plays in the kidney see Pluznick *et al*.[106]

BK_{Ca} dysfunction has also been linked to several pathophysiological conditions manifested in diabetes. While previous studies showed that overactive bladder was linked to decreased expression levels of α and β_1 subunits,[99] increased expression of α and decreased expression of β_1 has been observed in diabetic rat bladder smooth muscle, also leading to overactive bladder.[107] Recently, Yi *et al.* hypothesized a role for the protein MuRF1, a protein known to be involved in heart disease, in the proteolytic breakdown of β_1 and eventual diabetic vasculopathy, after observing decreased β_1 expression when MuRF1

expression increased in diabetic mouse vascular smooth muscle.[108] When BK_{Ca} α was blocked using iberiotoxin, which is discussed in section 1.4, a 70% increase in insulin secretion was observed in human pancreatic β islet cells[67]. In addition, increase in insulin was shown to increase BK_{Ca} α expression[109] suggesting a role for BK_{Ca} in regulating insulin secretion.

Finally, the molecular mechanisms by which obesity affects BK_{Ca} function are beginning to emerge. Studies on Zucker diabetic fatty rats, an obese diabetic rat model organism, have shown reduced activation of BK_{Ca} compared to non-fatty rats[110] and decreased activation of BK_{Ca} by arachidonic acid[111] in smooth muscle cells isolated from the arteries of Zucker diabetic fatty rats. Also, β_1 expression in obese rats was decreased in cremaster muscle arterioles, but increased in middle-cerebral arteries with no change in α subunit expression compared to non-obese controls.[112]

1.4 Channel Pharmacology and Therapeutic Potential

With the physiological significance of the BK_{Ca} channel established and the channel involvement in many pathological conditions coming to light, the value of the BK_{Ca} channel α and β subunits as a pharmacological target has received increasing attention. Many compounds have been discovered, which are discussed in more detail in section 1.4, that either activate or block channel activity. However, there have been no successful clinical trials on BK_{Ca} channel α or β subunit modulators so far. The lack of therapeutic agents targeting BK_{Ca} could be due to an absence of specificity or potency of compounds currently available. Most significantly, the scarcity of high resolution structural data contributes to poorly characterized binding sites for most ligands, impeding

efforts to develop alternatives ligands with improved potency and specificity. This section will focus on the compounds and peptides that target BK_{Ca} α or β subunits and what is known, if anything, about their binding sites and effects on channel behavior.

There are two peptide toxins known to target BK_{Ca} α subunits: charybdotoxin[113] and iberiotoxin.[114] They are both capable of inhibiting channel function at nanomolar concentrations by binding to the channel extracellular surface.[113] Moreover, a solution structure for unbound charybdotoxin has been reported.[115] While channels coexpressed with β_1 have the same affinity for charybdotoxin as α alone,[44] channels containing β_2 or β_3 show a 30-fold decrease in affinity,[44,47] while channels containing β_4 have a 1,000-fold decrease in affinity.[116] The difference in affinity in β_4 subunit containing channels is derived from the presence of several basic residues in the extracellular loop that prevent binding of charybdotoxin.[117] While charybdotoxin binding affinity for $\alpha+\beta_1$ channels remains the same, β_1 causes a 7-fold decrease in the dissociation rate constant (K_d),[118] an effect that seems to be associated with four residues in the β_1 extracellular loop.[119] Iberiotoxin differs from charybdotoxin in that it is specific for BK_{Ca} channel α subunits and has a K_d of 1 nm, 10 times lower than that of charybdotoxin, although both toxins appear to work by a similar mechanism that involves blocking the K^+ pore.[114]

Several small molecule BK_{Ca} channel modulators have also been reported. A series of small molecules derived from fungal toxins were shown to selectively block BK_{Ca} in bovine aortic smooth muscle cells.[120] Conversely, the

benzimidazolone series of compounds provided several potent channel openers including NS 004, NS 1609, and NS 1619 (Figure 2). NS 1619 was the first of these compounds found to evoke hyperpolarization of aortic smooth muscle cells[121] and rat neurons via activation of BK_{Ca} channels.[122] NS 1609 was later shown to open BK_{Ca} channels leading to hyperpolarization through channel opening in rat and human bladder smooth muscle cells in lower concentrations.[123] However, NS 1609 blocks channel opening when higher concentrations were applied in guinea pig bladder smooth muscle cells while NS 004 did not.[124,125] Recently, the mechanism of this class of channel openers was shown to require the linker between S6 and RCK1 of the α subunit, suggesting they elicit effects through the CTD.[126] Additionally, inhalation of NS 1619 was shown to reduce blood pressure in the right ventricle and improve blood oxygenation in a rat model.[127] However, Bristol Meyers Squibb (BMS) 204352 (Figure 2), a compound developed from the original benzimidazolone series, failed to show decreased neuronal cell death in clinical trails for the treatment of acute ischemic stroke,[128] which involves large influxes of Ca^{2+} leading to neuronal cell death, although it was shown to be a potent and selective α subunit channel opener.[129] Interestingly, NS 1619 is the only one in this series of compounds that does not effect other types of ion channels.[121] Although this series of compounds has seen some exploration for therapeutic development, the lack of specificity for BK_{Ca} channels and a lack of tissue specificity due to their action on the α subunit make these compounds less than ideal for therapeutic development.

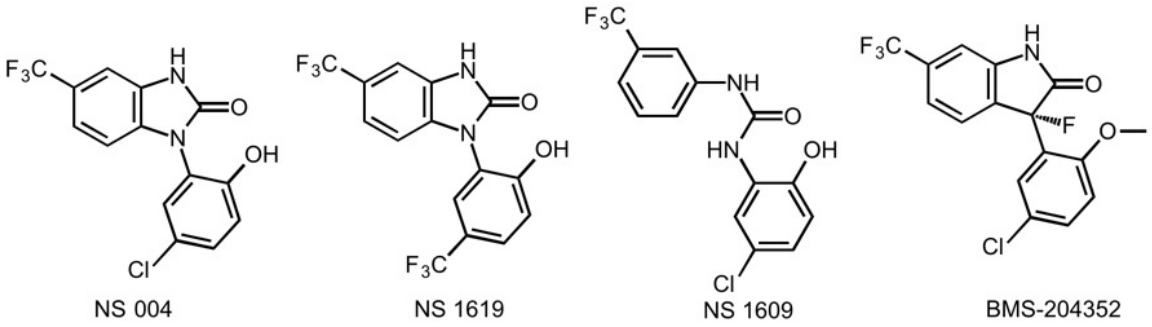


Figure 2. BK_{Ca} Channel Openers of the Benzimidazolone Analog Series

Bile acids (Figure 3) have also been shown to be micromolar activators of BK_{Ca} channels.[130] The cis stereochemistry of the A-B ring junction in the bile acid steroid backbone, a carboxylate functional group on the lateral chain, and a single hydroxyl with R stereochemistry on carbon 3 of the A ring were all shown to be important structural features required for channel activation.[130] It was later shown that the bile acids required the β_1 subunit for activation[131] with a putative binding site in the second transmembrane domain.[132] Moreover, lithocholic acid, the most potent bile acid activator of BK_{Ca} channels, had no effect on channels coexpressed with β_2, β_3, and β_4.[133] Later, the bile acid interaction site was shown to involve T169 in TM2 of β_1[65] and further structure activity relationship (SAR) studies refined the understanding of bile acid structural requirements.[134] However, these compounds are not ideal therapeutics for chronic disease treatment because of their steroid backbone, which makes them potential ligands of cytosolic and nuclear steroid receptors and thus, facilitates their wide involvement in several physiological processes and transcriptional regulation. For a review on bile acid biogenesis and function see Chiang 2013.[135] Finally,

cholesterol has been shown in multiple studies to decrease channel opening events.[136,137] The binding site of cholesterol was identified by our experimental collaborators at the University of Tennessee Health Science Center as a series of cholesterol amino acid consensus (CRAC) motifs in the CTD of the α subunit.[137] CRAC motifs have previously been shown to effect other membrane protein ion channels,[138] but the mechanism of the interaction between the amino acids in the motif and cholesterol are not known. Chapter 2 outlines computational experiments used to better understand the interaction between cholesterol and the protein, as well as explain the experimental findings of our collaborators.

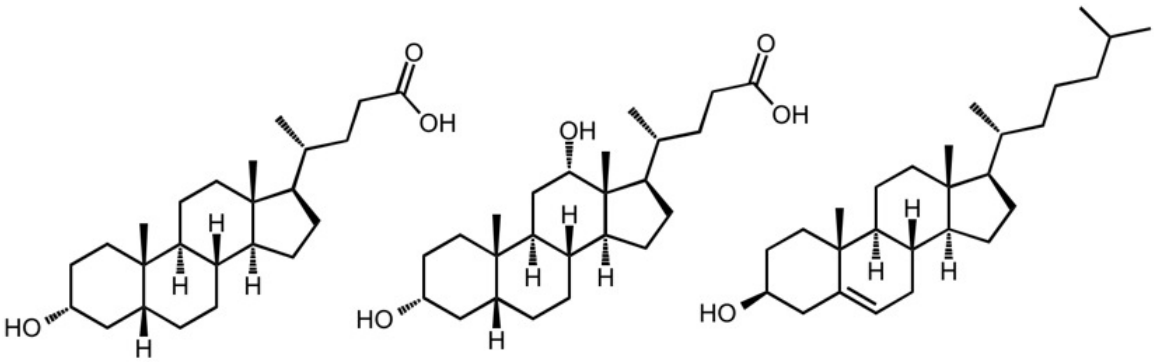


Figure 3. Examples of Bile Acids and Cholesterol that Modulate BK_{Ca}

There are two natural compounds obtained from herbs that activate BK_{Ca} channels. The first, dehydrosoyasaponin-1 (DHS-1, Figure 4), was reported to activate BK_{Ca} α subunits in 1993[139] and later shown to require the β_1 subunit to modulate the channel.[39] When applied to the inner artificial membrane leaflet, the compound reversibly increases the probability of the channel being open in patch

clamp experiments,[140] but due to the large polar sugar moiety preventing the compound from partitioning through the membrane, this compound does not produce an effect when applied to the outer membrane. Interestingly, a database search based on a pharmacophore derived from bile acids returned a related compound, 3-hydroxyolean-12-en-30-oic acid (3-HENA, Figure 4), that was shown to activate BK_{Ca} through the β_1 cholane steroid-recognition site with a greater increase in channel open probability and duration of effect than lithocholic acid.[141] Three-HENA was able to induce cerebrovascular dilation *in vitro and in vivo* through BK_{Ca} channels containing α and β_1 subunits, which underscores drug access to the channel from circulation and the extracellular side of the cell membrane.[141] While both DHS-1 and 3-HENA are triterpenoid compounds, 3-HENA lacks the sugar moiety thus allowing it to partition through membranes more readily than DHS-1. However, the hydrophobic nature of the compound would also present challenges in bioavailability if used as an oral therapeutic since it would not readily leave cell membranes and therefore would not reach the circulatory system in sufficient quantities if given orally. The other naturally derived compound, called MaxiKdiol (Figure 4), activates α subunits with or without β subunits present.[142]

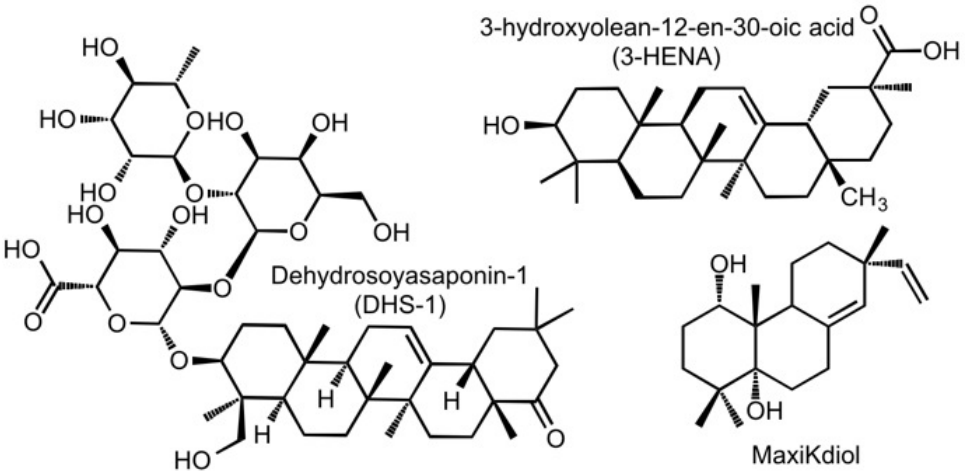


Figure 4. Plant Derived Compounds that Activate BK_{Ca}
DHS-1 and MaxiKdiol are both derived from herbs. DHS-1 and 3-HENA are both triterpenoid compounds, but 3-HENA was chemically modified after purification and is therefore semi-synthetic.

An interesting group of compounds that modulate BK_{Ca} are the estrogens, specifically estradiol, and estrogen mimetic agents (Figure 5) as well as anti-estrogens. Seventeen-β estradiol was first shown to activate BK_{Ca}, as well as voltage-gated Ca^{2+} channels, from the extracellular side of the membrane by Valverde *et al*[143] and was later shown to alter redox function in macrophages by activating BK_{Ca}.[144] Later, the N-terminal residues of the α subunit were shown to be necessary for the effect of estradiol on β_1 subunits.[15] Tamoxifen (Figure 5), an estrogenic compound, also induced activation of BK_{Ca} but required β_1 subunits;[145] both estrogen and tamoxifen were ineffective at inducing vasodilation in β_1 knockout mouse smooth muscle cells.[146] However, it has also been shown that an α subunit is sufficient for channel sensitivity to estradiol, while co-expression with the β_1 subunit increases the apparent binding affinity of estradiol.[147] The anti-estrogen compound fulvestrant[148] and bisphenol A, a compound developed as an estrogen mimetic that is now used in plastics,[149] have been shown to

produce the same effect on BK_{Ca} as estrogen. However, no interaction site has been established for estrogen or any of the estrogen mimetic compounds mentioned here. Without an interaction site and due to the many biological targets of estrogen hormones, pursuing therapeutics that mimic estrogen could have a wide and undesirable side effect profile.

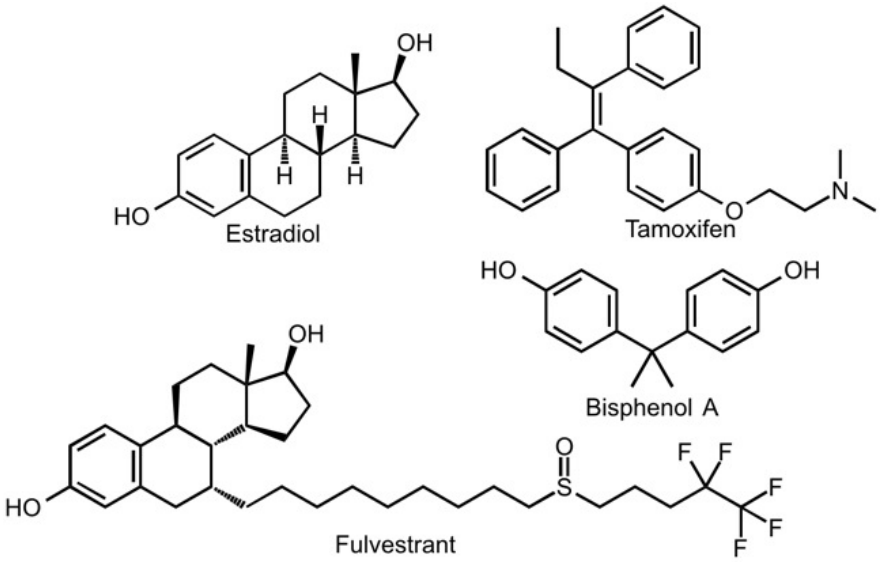


Figure 5. Estrogen, Estrogen Mimetics, and Anti-estrogen Compounds

Although there are many modulators of BK_{Ca}, including small molecules and peptides, very few of these compounds have known binding or interaction sites. Furthermore, only a small number of compounds (bile acids, and 3-HENA) are known to exert their effects on BK_{Ca} through a defined interaction site on the β_1 subunit. With a lack of defined binding sites, design and optimization of novel activators poses a significant challenge. Information about existing ligands can be useful for database mining, similarity searching, and pharmacophore searching, but structural data of the biological target would greatly enhance the

understanding of and design of potential new therapeutics. For an extensive review of BK_{Ca} channel pharmacology see Nardi and Olesen 2008.[150]

1.5 Molecular Dynamics to Study Biomolecular Interactions

While experimental techniques like alanine scanning mutagenesis can be useful for determining protein-ligand binding sites, these studies do not reveal the dynamic interactions between the protein and ligand. Conversely, molecular dynamics cannot be used to easily identify ligand binding sites due to limitations with computational resources and the time required for simulations. The combined application of experimental and computational approaches can be used to gain insight into protein-ligand interactions. Molecular dynamics (MD) simulations have been used to study proteins in a dynamic environment versus the static picture provided by X-ray crystal structures or solution and solid state NMR structures. Mathematical models of the dynamic multi-conformational properties of proteins first appeared in 1985[151] and have been reviewed extensively.[152,153] Although this is not an exhaustive list, MD simulations have been used to demonstrate how the dynamic structure of the protein can control ligand diffusion into active sites or binding sites[154,155] and modulate receptor affinity,[156] to understand the effects of protein mutagenesis on ligand interaction,[137,157] and even to elucidate the inhibition of influenza by mammalian proteins using large-scale simulations.[158] These different studies exemplify the vast applicability of MD in understanding complex biological macromolecules, their dynamic fluctuations, and their interactions with ligands or other macromolecules, but an understanding of how MD works and the limitations of

the simulations is required to appreciate, understand, and correctly implement the technique.

Although the concept of molecular dynamics was developed in 1959[159] for use in theoretical physics, it has expanded, due to the advancement of microprocessors, to study large complex systems of biomolecules. In their most basic form, MD simulations use the Hamiltonian operator of every atom in the system treated classically and solves for the forces acting on the atom by integrating Newton's equations of motion.[160] For every time step in a simulation, the equations of motion are integrated and new velocities and positions for each atom are acquired.[161] Every MD program must contain a model for interaction of all atoms or molecules, an integrator to solve atomic positions and velocities over the time steps of the simulation, and a statistical ensemble that controls the thermodynamic properties of the simulation, including the number of particles in the system, temperature, pressure, and volume.[161] Due to the complexity of these calculations, the size of a system simulated and the length of time simulated are inherently limited. Newer parallel processing technologies, like NVIDIA CUDA, have greatly expanded the feasible number of atoms in a simulation and the length of the simulation, while also reducing the number of computing hours needed to complete the simulation. This brings simulations within reach that were previously not possible because of scale. The picosecond to microsecond time scale now possible with MD allows for the observation of molecular collisions and bond rotation and vibration, but cannot simulate the movement of electrons due to the use of classical mechanics.[161] The use of

classical mechanics is a limitation that prevents the simulation of bond forming or bond breaking using traditional MD, but may be overcome using hybrid quantum mechanics/molecular mechanics methods.

A major part of any MD simulation is the use of a force field to parameterize the molecules in the system. The force field in any simulation has an impact on the accuracy of the simulations performed, so care must be taken when choosing a force field to select one with appropriate parameters for the system being studied. Force fields must contain potential energy parameters for van der Waals interactions, bond stretching and bending, and torsional angles for all atoms and bond types in the system.[161] These parameters must be fitted to reproduce empirical or *ab initio* calculation results. Additionally, both short- and long-range interactions must be accounted for and are typically treated separately. In AMBER 10,[162] the software package used for simulations in Chapter 2, short-range interactions are approximated using a Lennard-Jones (LJ) potential[163] and long-range interactions are approximated using the Particle Mesh Ewald (PME)[164] technique, while the ff99SB force field[165] uses an empirically derived potential. The PME technique has been shown to overcome shortcomings of other long-range approximation methods to reach agreement with experimental results[166] and the LJ potential is implemented because of its mathematical simplicity and accuracy in approximations.[161]

In conclusion, the use of molecular dynamics to explore protein-ligand interactions is a valuable tool to explain experimentally obtained results. While there are limitations on the applicability and what non-bonded interactions can be

accounted for because of approximations used to overcome computational limits, proper knowledge of MD and those limitations can yield high-value data to explain molecular behavior. For an extensive explanations of all equations, potentials, integrators, ensembles, and parallelization methods see Sutmann.[161]

1.6 Ligand-Based Pharmacophore Modeling for Therapeutic Lead Compound Discovery

Although it has been used in many ways and to mean many different things, the term pharmacophore is defined by the International Union of Pure and Applied Chemistry as "the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response" with pharmacophore descriptors being the points that define a pharmacophore.[167] While three dimensional (3D) pharmacophores were described before computers were used for 3D modeling,[168] the first semi-automated pharmacophore modeling was described by Mayer *et* al in 1987.[169] After that, a new era of computer aided drug design (CADD) made building models and searching molecular conformational databases feasible. Now, many programs exist to develop pharmacophore models using increasingly sophisticated algorithms that allow for the generation and testing of models. Models can be developed based on the structure of known ligands for a particular target, called "ligand-based modeling," or can be generated from a known protein target structure, called "structure-based modeling". With very little structural information available for $BK_{Ca,}$ a lack of well-characterized binding sites, and no structural data available for the β_1 subunit, the pharmacophore model discussed in Chapter 3 was developed using

a ligand-based method. Models of both types can be assessed for quality using the Güner and Henry method,[170] which involves testing the predictive value of models against a set of compounds with known activity at the target protein. Validated models can then be used for database searching to find novel compounds that can be screened against the target protein *in vitro.* For an extensive review of the current state of pharmacophore modeling and future challenges see Yang 2010,[171] Wolber 2008,[172] or Langer 2006.[173]

Since structural data is not currently available for the β_1 subunit, but a cholane steroid-recognition site has been established, Chapter 3 of this dissertation describes the advancements in developing a ligand-based pharmacophore model for this interaction site. By utilizing the insight gained from previous structure activity relationship studies, models were developed that allow for rapid screening of available conformational databases to find compounds that can interact with the cholane steroid site in β_1. With an end goal of discovering non-steroidal compounds that could be used to treat chronic conditions such as hypertension, the development of models for this interaction site make this goal attainable, although experimental validation of models is still needed.

1.7 Transmembrane Protein Expression and Purification

Transmembrane (TM) proteins represent a vast number of proteins encoded by the human genome with 37% of all open reading frames encoding proteins with at least one TM segment.[174] Examples of transmembrane proteins are protein receptors that bind ligands on the outside of the cell and transmit a signal inside the cell, like G-protein coupled receptors, and transporter proteins that move molecules, ions, or peptides into or out of the cell, like the BK_{Ca}

channel. While TM proteins play a significant role in maintaining cellular homeostasis and receiving extracellular signals to coordinate cellular functions that make multi-cellular organisms possible, very few atomic resolution structures of TM proteins exist. Although the first atomic resolution membrane protein structure was published in 1985,[175] only 464 (http://blanco.biomol.uci.edu -/mpstruc/) of the 99,122 structures deposited in the Protein Data Bank[176] (both websites were accessed April 5th, 2014) are unique membrane protein structures, which demonstrates the difficulty of membrane protein structural characterization. The deficit of structural data about transmembrane proteins combined with the physiological importance of the BK_{Ca} channel and the tissue specific expression of the β_1 subunit outlined in section 1.3, underscores the desire to pursue structural characterization of the β_1 subunit. While pieces of the α subunit structure and small fragments of some β subunits have been reported (see section 1.1), neither a complete structure of the channel nor any of the β subunits has yet to be reported.

The reason for the relatively small number of available structures stems from the difficulty of maintaining a membrane-like environment and preventing amorphous protein aggregation from large spans of non-polar amino acids.[177] In living cells, this process is handled by translocon cellular machinery that traffics membrane proteins from the cytosol where they are produced to the cell membrane,[178] but after expression of a target protein in a host organism the protein must be extracted, purified, and stabilized for structural studies outside of the native cell membrane environment. While different systems have been

employed, which will be discussed in more detail in Chapter 4, there is currently no way to determine which system will work with a particular protein target, making membrane protein purification and stabilization a laborious process. Additionally, *Escherichia coli,* the host organism chosen for protein expression in Chapter 4, presents more challenges when expressing eukaryotic membrane proteins with prokaryotic cellular machinery, including inclusion body formation and lacking eukaryotic mechanisms for post-translational modification of proteins.

Membrane proteins are divided into the β barrel class and the α helical class[179] with the BK_{Ca} channel belonging to the latter. While many β barrel membrane proteins have been successfully denatured and refolded from inclusion bodies, it is more difficult with α helical proteins. In bacteria, the formation of inclusion bodies is caused by many factors, but the biggest contributors are the stress response from the cell and overwhelming the translocon machinery leading to truncated or misfolded proteins aggregated in the cytoplasm.[180,181] Overcoming these problems has been achieved using fusion proteins that enhance solubility or target the protein for membrane insertion, weaker T7 promoters for expression by T7 RNA polymerase that slow expression rates, and using various *E. coli* cell lines optimized to express toxic or membrane proteins.[182] While some of these methods were utilized in this study, the lack of certain post-translational modifications utilized in eukaryotic cells that are not present in *E. coli* could result in problems with protein structure and function.

Post-translation modifications, used in eukaryotic cells to direct proteins to their destination or assist in folding and structure, do not occur to the same

extent in *E. coli*. As a result of bacteria lacking asparagine-linked post-translational modification, problems with expression of β_1 in *E. coli* could arise from the presence of two N-linked glycosylation sites in the protein.[183] Since glycosylation often plays a role in maintaining proper protein structure in eukaryotic cells,[184] misfolded or non-native protein structure could result from using a bacterial expression system. Previous studies demonstrated that enzymatic deglycosylation of β_1 has a significant impact on channel properties causing an increase in open probability and mean open time of channels in patch-clamp experiments.[185] However, the modulatory role of β_1 was not eliminated, suggesting that some structural features were retained in the non-glycosylated protein. It is possible, however, that β_1 would not fold to the proper functional form if the sugars had been absent during expression. Because of this, expression in *E. coli* could lead to non-native folding and potential aggregation or, if a structure were obtained, a structure with altered modulatory function relative to wild type β_1. However, any structural insights gained would greatly improve upon what is currently known.

The goal of Chapter 4 was to express and purify β_1 protein in quantities sufficient for structural characterization using solution nuclear magnetic resonance (NMR) spectroscopy. While the protein was successfully expressed and purified, assessing the functionality of the protein remains a challenge necessary to overcome to justify the time and expense of NMR experiments. Although functionality could not be demonstrated, successful expression and purification brings structural characterization of β_1 within reach.

Chapter 2

Evaluation of the Interactions of Cholesterol with the BK_{Ca} Channel Cytosolic Domain Using Molecular Dynamics

2.1 Introduction

Biological lipids are involved in many physiologically important roles including the formation of cell membranes, cell signaling events, energy storage, and compartmentalization of cellular functions. Cholesterol (CLR) has been shown to play an important role in modulation of membrane protein functions through the cholesterol recognition amino acid consensus (CRAC) motif. This motif is loosely defined by the amino acid sequence L/V - X_{1-5} - Y - X_{1-5} - R/K (where X is any amino acid and "/" means or).[186] Furthermore, CLR has been shown to alter the activity of members of every ion channel family.[138] The action of cholesterol on these ion channels has been proposed to rely on protein-CLR interaction, alterations of membrane properties that subsequently alter ion channel activity, or modification of protein-protein interactions by CLR.[138]

The modulation of the BK_{Ca} channel by CLR was first shown in 1989 when the ionic current of the channel was shown to decrease in the presence of CLR.[136] At the time, membrane fluidity was hypothesized to cause the decrease in current, but more recently the effect of CLR was shown to involve several CRAC domains in the cytosolic domain of the protein.[187] In the α subunit, ten CRAC sequences were identified with CRAC4-10 in the CTD portion of the protein.[187] The CTD CRAC sequences were shown to be responsible for the CLR sensitivity of the protein through mutagenesis studies[187] and are discussed below. Additionally, experiments showed that the enantiomer of CLR does not

alter channel currents[188,189] supporting that the sensitivity of BK_{Ca} to cholesterol is likely due to a stereospecific interaction with the protein and not a result of altered membrane dynamics.

In experiments on BK_{Ca} CRAC domains performed by our collaborators at the University of Tennessee Health Science Center, several truncations were made to elucidate which CRAC motifs identified from the primary sequence contributed to CLR sensitivity, as well as several mutations within CRAC motifs to determine the role of different residues in CLR sensitivity. The first truncation was immediately after the S6 transmembrane segment (Figure 1) and resulted in a channel insensitive to CLR despite the channel remaining functionally active.[137] This indicates the CRAC motifs found in the transmembrane segments (CRAC 1-3) were not sufficient for CLR sensitivity in BK_{Ca} α subunits.[137] Additional truncations were made in the CTD portion of the protein after CRAC4, CRAC6, and CRAC8 to determine to contribution of the cytosolic CRAC motifs to CLR sensitivity. CLR sensitivity was successively decreased compared to wild type (WT) protein with truncations after CRAC6 and CRAC8.[137] Truncation after CRAC4 produced a protein more sensitive to CLR than WT, although the reason for this could not be determined.[137]

The increased CLR sensitivity of the protein truncated after CRAC4 led to mutagenesis studies of CRAC4 residues V444, Y450, and K453 to produce V444A, Y450F, and K453A mutants.[137] The V444A mutation produced a protein that was resistant to CLR inhibition, but not completely insensitive.[137] The Y450F and K453A mutations completely abolished CLR sensitivity, underscoring their

importance in CLR recognition in the CRAC4 motif.[137] The results of the Y450F and K453A mutations drove the molecular dynamics simulations of CLR in this study.

Initial molecular dynamics simulations spanning 5 ns of simulation time were performed to elucidate differences in CLR behavior between wild type (WT) BK_{Ca} CTD and the proteins truncated after CRAC4, as well as two mutants, Y450F and K453A, to mimic the mutants examined experimentally. These simulations showed differences in time spent by CLR on the CRAC domain in the various constructs,[187] but were only 5 ns in length, so longer simulations were needed to fully discern patterns or differences in molecular behavior between constructs.

2.2 Methods

The computational methods described in sections 2.2.1and 2.2.2 were originally published in the Journal of Biological Chemistry. Singh, A. K.; McMillan, J.; Bukiya, A. N.; Burton, B.; Parrill, A. L.; Dopico, A. M. Multiple cholesterol recognition/interaction amino acid consensus (CRAC) motifs in the cytosolic C tail of the slo1 subunit determine cholesterol sensitivity of Ca^{2+}- and voltage-gated K^+ (BK) channels. *Journal of Biological Chemistry*. 2012; 287:20509-21. © The American Society for Biochemistry and Molecular Biology. The methods presented are the same, but simulations were extended to 200 ns and descriptions of methods were altered to reflect terminology used in this dissertation.

2.2.1 Protein Structure Preparation

To help elucidate the nature of the interactions between CLR and the BK_{Ca} α subunit, simulations of the full cytosolic tail domain (CTD, residues 331-1118) and the truncated CTD (residues 331-456; referred to as trCTD) polypeptides were compared with that of mutants Y450F and K453A (simulated as truncated proteins and referred to as Y450F or K453A). The starting structures of the channel protein were based on the crystallized portion of the BK_{Ca} CTD (Protein Data Bank[176] entry 3NAF[53]). The crystallized sequence was aligned with its full sequence from the UniProt database[190] (accession number Q12791), and Molecular Operating Environment (MOE 2010.10)[191] software was used to create a three dimensional homology model for the missing loops from residues 617 to 716, 726 to 742, and 893 to 930 in the PDB crystal structure. The model was geometry-optimized using ff99SB,[165] a forcefield parameterized for proteins, to a root mean square gradient of 0.5 kcal/mol·Å . The truncated structures were made by removing all residues after Ile-456, which is consistent with the truncation made for patch clamp experiments on recombinant channel protein. Mutated structures for Y450F and K453A were based on the truncated structure and mutations were made in MOE 2010.10 before geometry optimization. The preparatory files for the CLR molecule were generated in antechamber[192] using semiempirical AM1-BCC charges,[162] which have been parameterized to reproduce HF/-31* RESP geometry and atom charges. Counterions were added via teLeap to neutralize the systems. A salt concentration of 53 mM was created using 50 K^+ and 50 Cl^- ions for CTD simulations and 20 K^+ and 20 Cl^- ions for the truncated CTD simulations. The systems were solvated with TIP3P water[193] with

15.0 Å padding on each side in a periodic box. CLR was manually placed above the CRAC4 motif with the sterol methyl groups facing the motif and the CLR hydroxyl placed in the vicinity of Lys-453 with the CLR hydrophobic tail near Tyr-450, as described in previous models.[186,194]

2.2.2 Molecular Dynamics

Atomistic molecular dynamics (MD) simulations were run using the AMBER 10 software package[162] on intracellular portions of the BK_{Ca} channel α subunit CTD and interacting CLR using the ff99SB[165] and gaff force fields,[195] which are optimized for proteins and organic molecules, respectively. Two rounds of minimizations were run: 1,000 steps (2 ns) with the protein and CLR fixed and 100,000 steps (200 ns) with the molecules unrestrained. A restrained warm-up MD was run using a time step of 2 fs at constant volume for 100 ps, with a restraint constant, or energy penalty for moving an atom, of 10.0 kcal/mol·Å^2, which is sufficient to prevent movement from the starting position during equilibration phases, on the protein and CLR molecules. During this simulation, the temperature was raised to 300 K using the Langevin thermostat with a collision frequency of 1.0 ps^{-1}. Water and ion densities were equilibrated via a restrained NPT ensemble (fixed number of particles N, pressure P, and temperature T) MD simulations (p = 1.0 bar) for 2 ns, using the same restraint, thermostat, and collision settings as during the warm-up. After 2 ns, the restraint was lifted, and production runs were recorded, using a time step of 2 fs. All MD simulations used the SHAKE algorithm[196,197] to constrain covalently bonded hydrogen atoms and the Particle Mesh Ewald method[164] to calculate long range

electrostatic interactions, using a cut-off distance of 12.0 Å. Histidines were represented as HIE (neutral charge: Nε-H, Nδ).

2.2.3 Molecular Dynamics Trajectory Analysis

The 200 ns trajectories obtained for each construct were analyzed using PTRAJ and CPPTRAJ[198] from AmberTools 13.[199] Visualization of molecular dynamics trajectories was accomplished using VMD 1.9.1[200] (http://www.ks.uiuc.edu/Research/vmd/). Images were rendered using the Tachyon ray tracing library[201] with secondary structure predicted using STRIDE.[202] Solvent exposed surface area calculations were completed using SURF[203] implemented in VMD.

2.3 Results and Discussion

Table 1 provides the residue numbers in the full BK_{Ca} sequence along with their residue numbers in simulation files. The CRAC motifs will be referred to by their number or BK_{Ca} residue numbers in this dissertation unless specified otherwise. The corresponding residue numbers are provided for context along with the sequence for each CRAC motif.

Table 1. CRAC Motif Residue Numbers in Human BK_{Ca} Protein and Corresponding Values in Simulation Files

CRAC Motifs	BK_{Ca} Residue Number	Simulation Residue Number	Amino Acid Sequence
CRAC4	444-453	114-123	VISIKNYHPK
CRAC5	461-466	131-136	LQYHNK
CRAC6	564-575	234-245	LMIAIEYKSANR
CRAC7	605-613	275-283	VKRAFFYCK
CRAC8	815-822/23	485-492/93	VGSIEYLK/R
CRAC9	1048-1060	718-730	LGDGGCYGDLFCK
CRAC10	1068-1076	738-746	LCFGIYRLR

2.3.1 Protein Stability During Simulations

The RMSD of the four different constructs (CTD, trCTD, K453A, and Y450F) were calculated for each 200 ns simulation to determine if protein stability remained constant over time. For each construct, RMSD calculations were performed only on protein residues and did not include CLR. Figure 6 shows the trCTD, and the two mutant simulations K453A and Y450F that were based on the trCTD construct, have stable RMSDs of 4-5 Å over the 200 ns simulation. The CTD construct had a less stable RMSD over time, but all four simulations remained 7-8 Å over 200 ns. Simulation 4 of the CTD construct reached an RMSD of 10 Å early in the simulation, which could be a problem with the overall stability of the protein. Calculations of per residue deviations were necessary to determine if the whole protein in the CTD simulation set was deviating or if there were sections of the protein that were causing the overall RMSD to remain higher than the trCTD and mutant simulations.

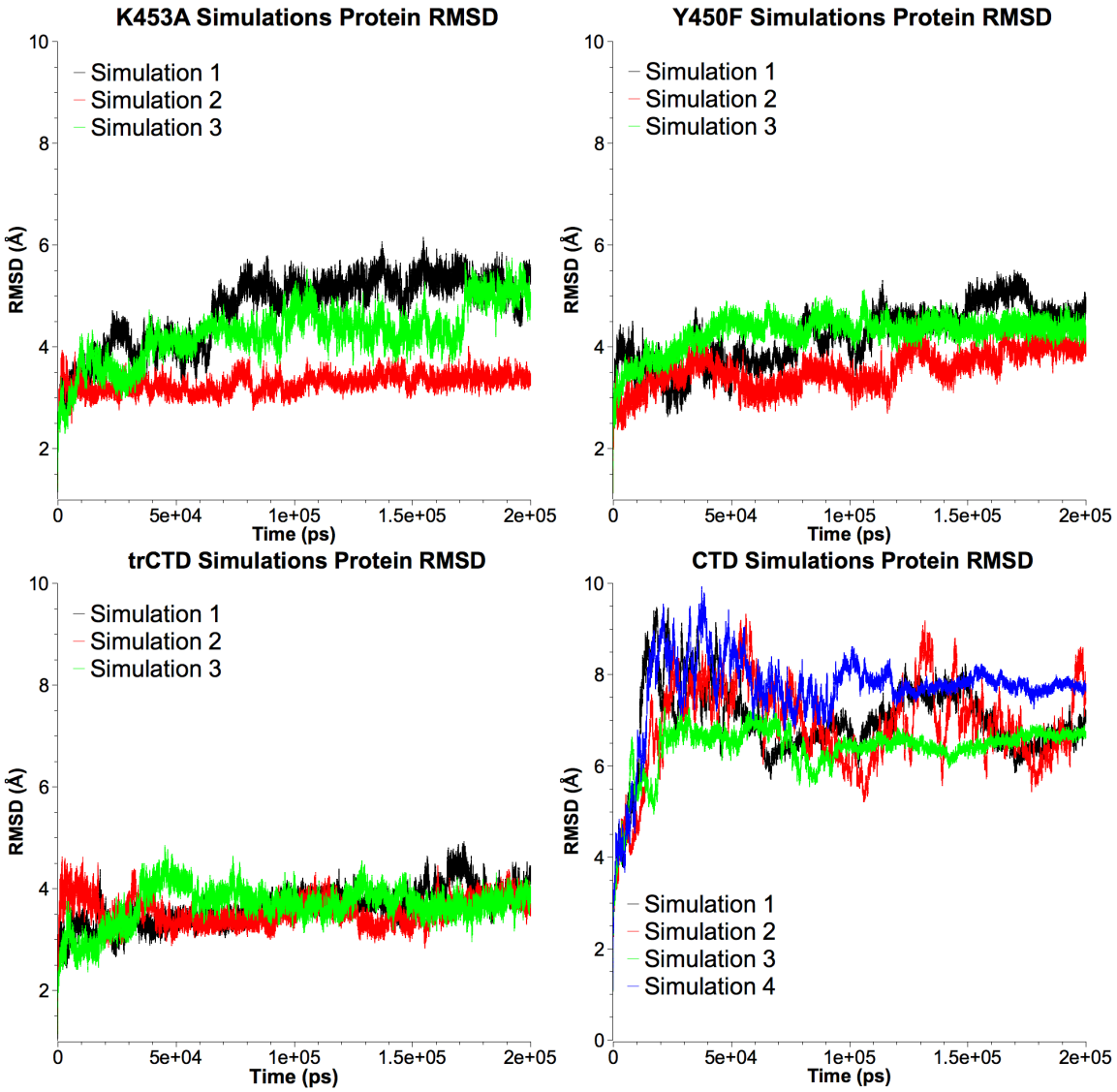
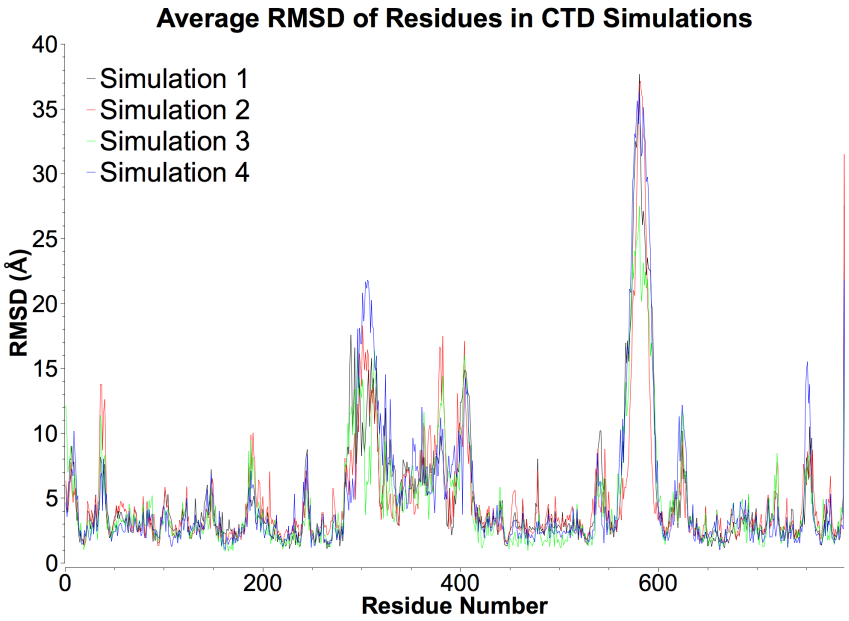


Figure 6. RMSD of Proteins in All Simulation Constructs
All three constructs based on the truncated protein (trCTD, K453A, and Y450F) achieved a stable RMSD around 4-5 Å. CTD simulations had higher RMSDs around 7-8 Å in all four simulations.

To examine the CTD simulations further, the RMSD was calculated for each residue for each of the four simulations over 200 ns. The per residue RMSD values are plotted in Figure 7. Residue numbers in simulation constructs are 1-787, but correspond to 331-1117 in the full protein sequence. Residues 287-387 in Figure 7 show a high RMSD. These residues in the simulation construct correspond to residues 617-717 in the full protein sequence, a portion of the protein following CRAC 7 and containing two of the amino acid segments missing in the crystal structure template, which were modeled using MOE as described in section 2.2.1. The second area in the simulation construct with abnormally high RMSD values (residues 563-600) also corresponds to a missing loop in the crystal structure that was modeled by MOE; residues 563-600 fall midway between CRAC 8 and 9. Additionally, residues at the amine and carboxy terminus of the protein have high RMSD values, as is expected since these residues are not anchored and have random coil secondary structure. Since these areas have high RMSDs throughout the simulations, the total RMSD for all residues (Figure 6) in CTD simulations is higher than in the trCTD, K453A, and Y450F simulations that contain only residues 1-126, all of which were resolved in the crystal structure used to develop the starting structure for simulations.

A



B

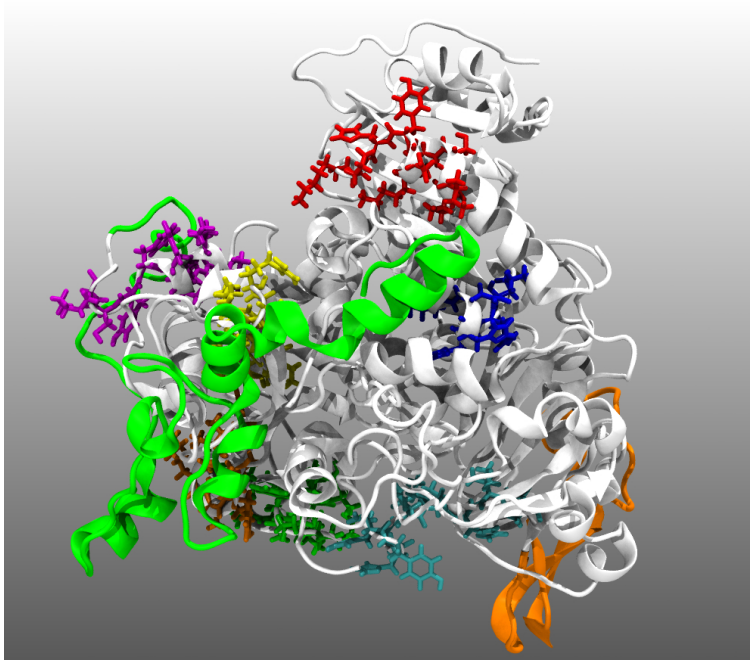


Figure 7. Average RMSD of Residues in CTD Simulations
A) Residues 1-787 in simulation constructs correspond to residue 331-1117 in the protein sequence. High RMSD values were calculated for residues found in sections missing from the crystal structure and modeled using MOE prior to simulation. B) Protein shown as ribbon structure. Residues 287 to 387 shown in green ribbon and residues 563 to 600 shown in orange ribbon. All CRAC4 (red), CRAC5 (blue), CRAC6 (yellow), CRAC7 (magenta), CRAC8 (teal), CRAC9 (green), CRAC10 (orange) residues shown as stick models

Figure 7B shows the high RMSD regions colored as a green (residues 287 to 387) and an orange (residues 563 to 600) ribbon structure. Due to the limited secondary structure of the orange region and the fact that it does not closely associate with the remainder of the protein, this part of the protein is highly mobile during simulation. The green region also associates very little with other parts of the protein, but has some secondary structural ordering in helices. However, these highly mobile regions do not contain CRAC motif residues (Figure 7B CRAC residues shown as blue stick models) and should have no impact on CLR behavior in CTD simulations since major protein conformational changes are not observed on this time scale in MD simulations. Because the protein remains stable over simulation time for each construct, the movement of CLR in each simulation can be evaluated without changes in protein conformation being a factor contributing to differences.

2.3.2 Cholesterol Interactions with CRAC Motifs

To determine if CLR exhibits variable dynamics in different simulated constructs, interatomic interactions within 5 Å of the H9 hydrogen atom in CLR (Figure 8) and any atom in the residues of the CRAC motifs (Table 1) were calculated. This atom was chosen in CLR because of the central location in the steroid nucleus. These interactions were calculated for CRAC4 residues in trCTD, K453A and Y450F simulations and CRAC4-CRAC10 residues for CTD simulations. Due to the CRAC4 motif being split between spatially distinct helix and random coil sections of the protein crystal structure, the CRAC4 contacts were separately calculated for residues 444 to 449 and 450 to 453 (Figure 9) and plotted together as discussed below. The plot for each construct below shows the

total number of interatomic contacts for each frame of the simulation, which is equivalent to 1 ps of simulation time.

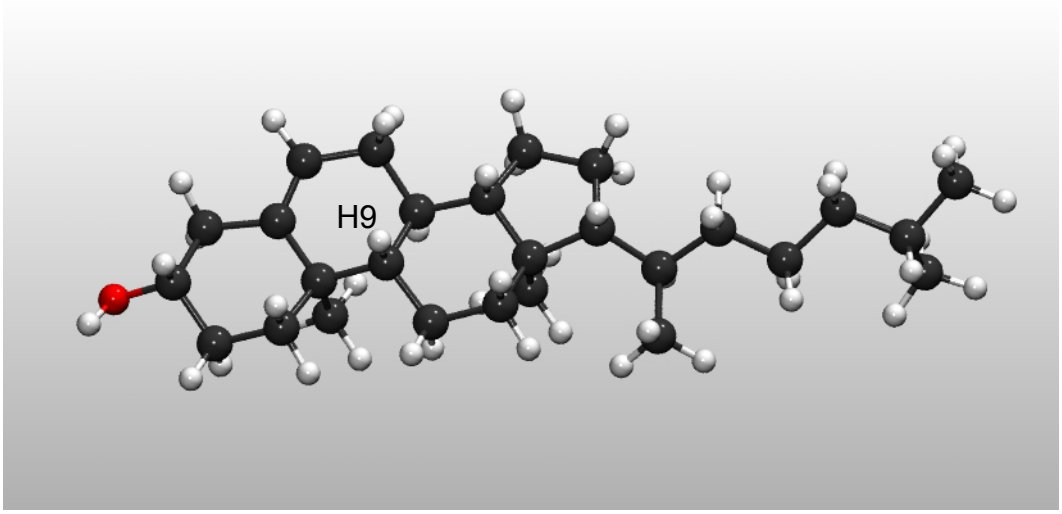


Figure 8. Hydrogen Atom in Cholesterol Used to Calculate Interatomic Interactions with CRAC Motifs
Interactions with CRAC motifs calculated within 5 Å of H9 atom in CLR for all simulations.

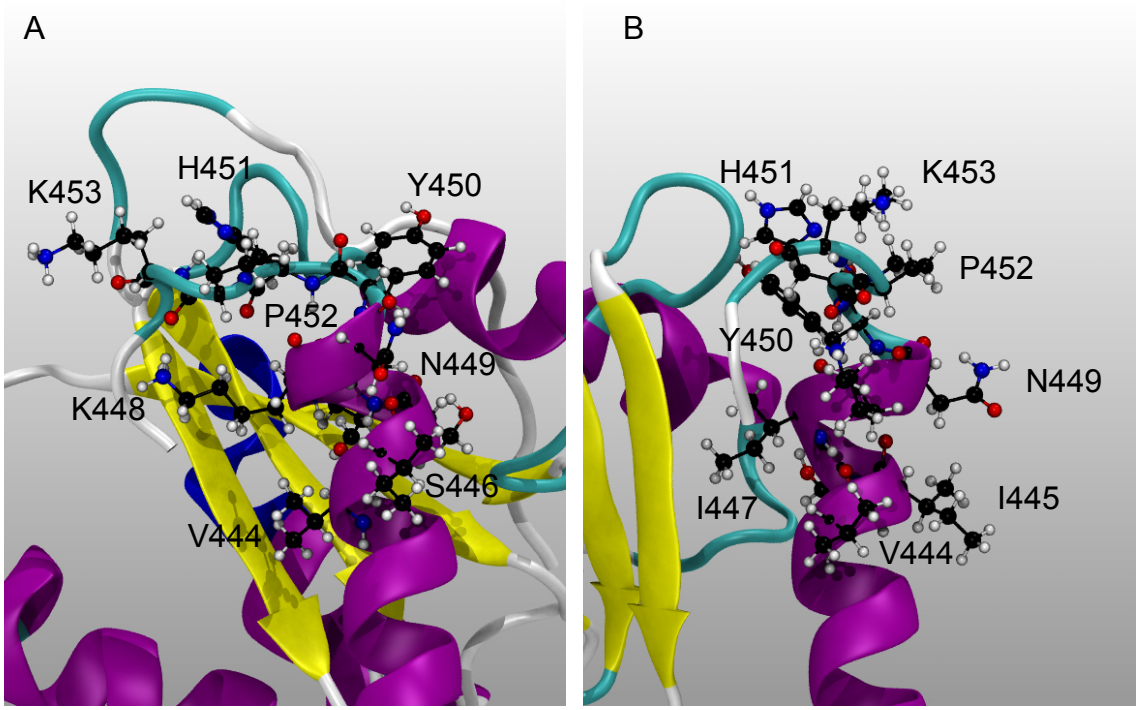


Figure 9. Residues of the CRAC4 Motif
A) Front view of CRAC4 motif with residues 450 to 453 in random coil (teal) and residues 444 to 449 in the helix (magenta). B) Side view of residues in CRAC4 motif. Not all residues are labeled in each image, but all residues are labeled in at least one image.

In trCTD simulations (Figure 10, Figure 11), the first and third simulations show CLR molecules remaining on or near CRAC4 residues 450 to 453 through 145 ns of simulation time. In simulation 1, an overlap region from 120 ns to 150 ns can be seen with interactions between CLR and both sets of CRAC4 residues, but only interactions with residues 444 to 449 occur after 150 ns. The CLR molecule in both simulations begins to transition from residues 450-453 to residues 444-449 after 120 ns in simulation 1 (Figure 11A) and after 160 ns in simulation 3 (Figure 11C). Simulation 2 shows different behavior for CLR compared to simulations 1 and 3. In the second simulation (Figure 11B), CLR starts on the CRAC4 motif and migrates away to another part of the protein not identified as a CLR recognition site and remains away from CRAC4 residues throughout simulation time, coming close enough to register a small number of interactions briefly around 170 ns. Additionally, the hydrogen bonds between the hydroxyl of CLR and the side chain of K453 are shown as green circles on each plot in Figure 10. In all three simulations, hydrogen bonds occur in high frequency during times where interatomic interactions are high (10 to 15 interactions per frame) or immediately preceding high interaction periods. This indicates the hydrogen bond between K453 and CLR may be important in CLR recognition or keeping CLR in contact with CRAC residues and is discussed in more detail in section 2.3.3.

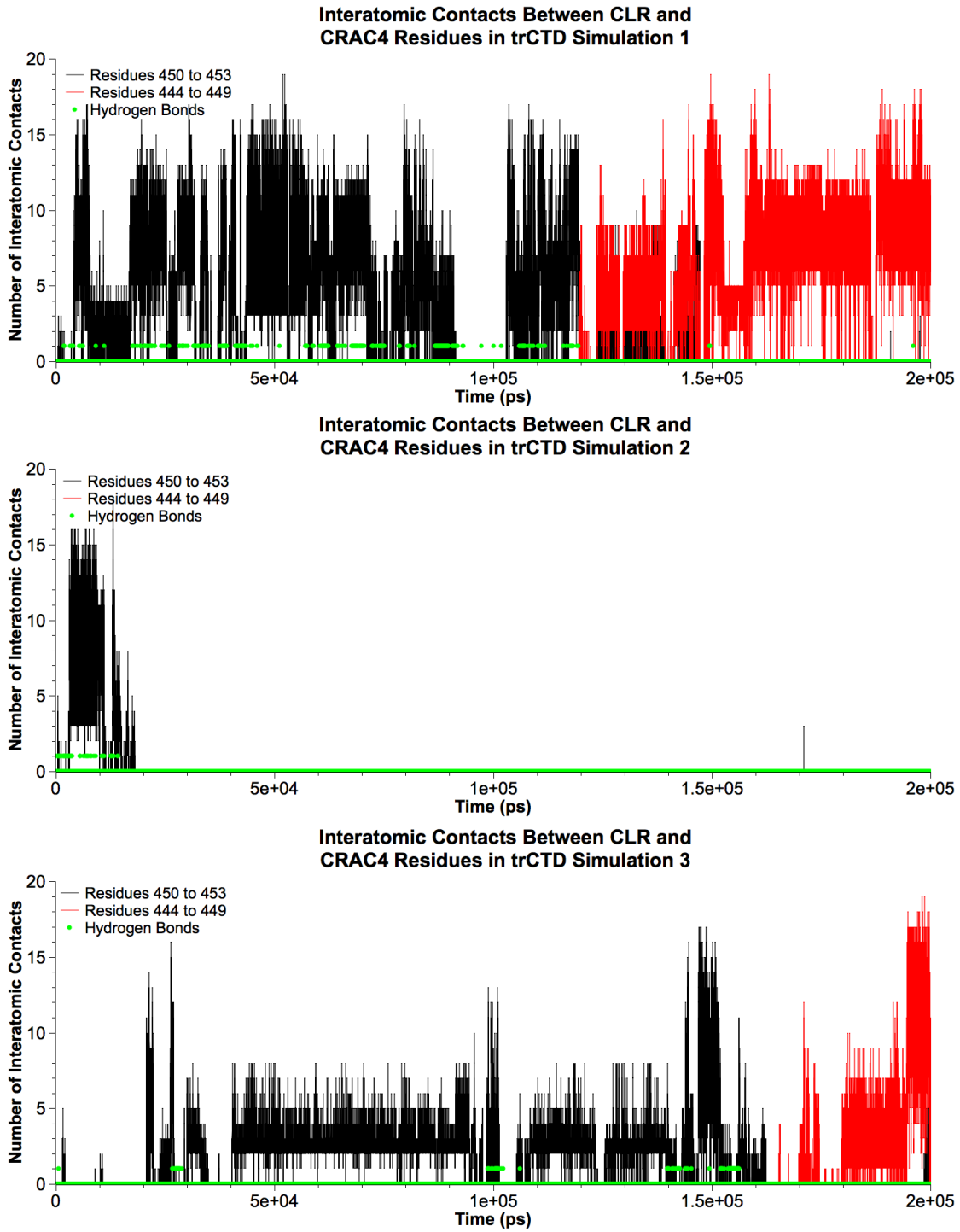


Figure 10. trCTD Simulations Cholesterol and CRAC4 Interatomic Contacts Contacts less than 5 Å between CLR H9 atom and residues in CRAC4 shown as total number of contacts per frame (1ps) of each simulation. Green circles represent points in the simulation where K453 hydrogen bonds with the hydroxyl of CLR.

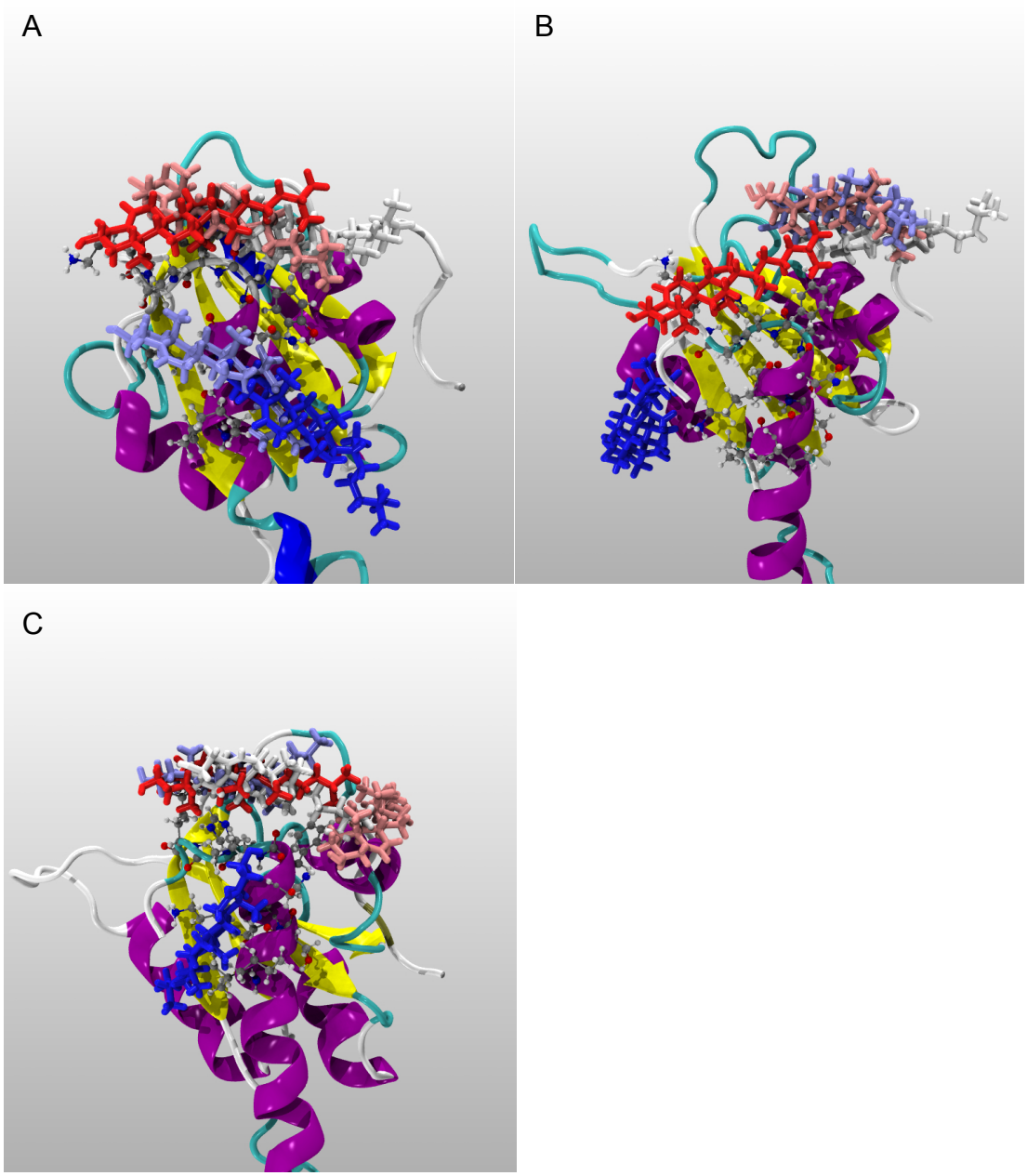


Figure 11. Positions of Cholesterol on CRAC4 Motif in trCTD Simulations
CLR (stick models) coordinates in 200 ns simulations shown every 50 ns. Time steps shown are 0 ns (red), 50 ns (pink), 100 ns (white), 150 ns (purple), and 200 ns (blue). Protein is shown as ribbon structure with CRAC4 residues as ball and stick models. A) Simulation 1 has CLR remain on the CRAC4 motif throughout the simulation. B) Simulation 2 shows CLR molecule starting on CRAC4 (red) and moving around the protein early in the simulation. C) Simulation 3 shows CLR remaining on CRAC4 throughout simulation time, but beginning to move away at 200 ns.

For K453A mutant simulations, the CLR molecules migrate quickly from residues 450-453 to the spatially distinct residues 444-449 in a helical section of the protein (Figure 12). Simulation 2 shows 5-15 interatomic interactions with residues 444 to 449 after 100 ns, indicating close proximity to several of the residues. This was confirmed by examining the positions of CLR throughout simulation time, where Figure 13B shows CLR in close proximity to residues in the helical portion of CRAC4 in the 100 ns, 150 ns, and 200 ns snapshots (colored as white, purple, and blue, respectively). The third simulation shows CLR moving away from all CRAC4 residues around 40 ns and moving to a separate part of the protein, but returning to CRAC4 residues 444 to 449 in the last 50 ns of simulation. The variable behavior of CLR on K453A simulations compared to trCTD simulations, where CLR stays on residues 450 to 453 through 150 ns of simulations 1 and 3, could be due to the K453A mutant lacking a side chain capable of hydrogen bonding with the hydroxyl of CLR. For comparison, the number of hydrogen bonds between the CLR hydroxyl and residue K453 are shown for the trCTD simulations in Figure 10 as green circles.

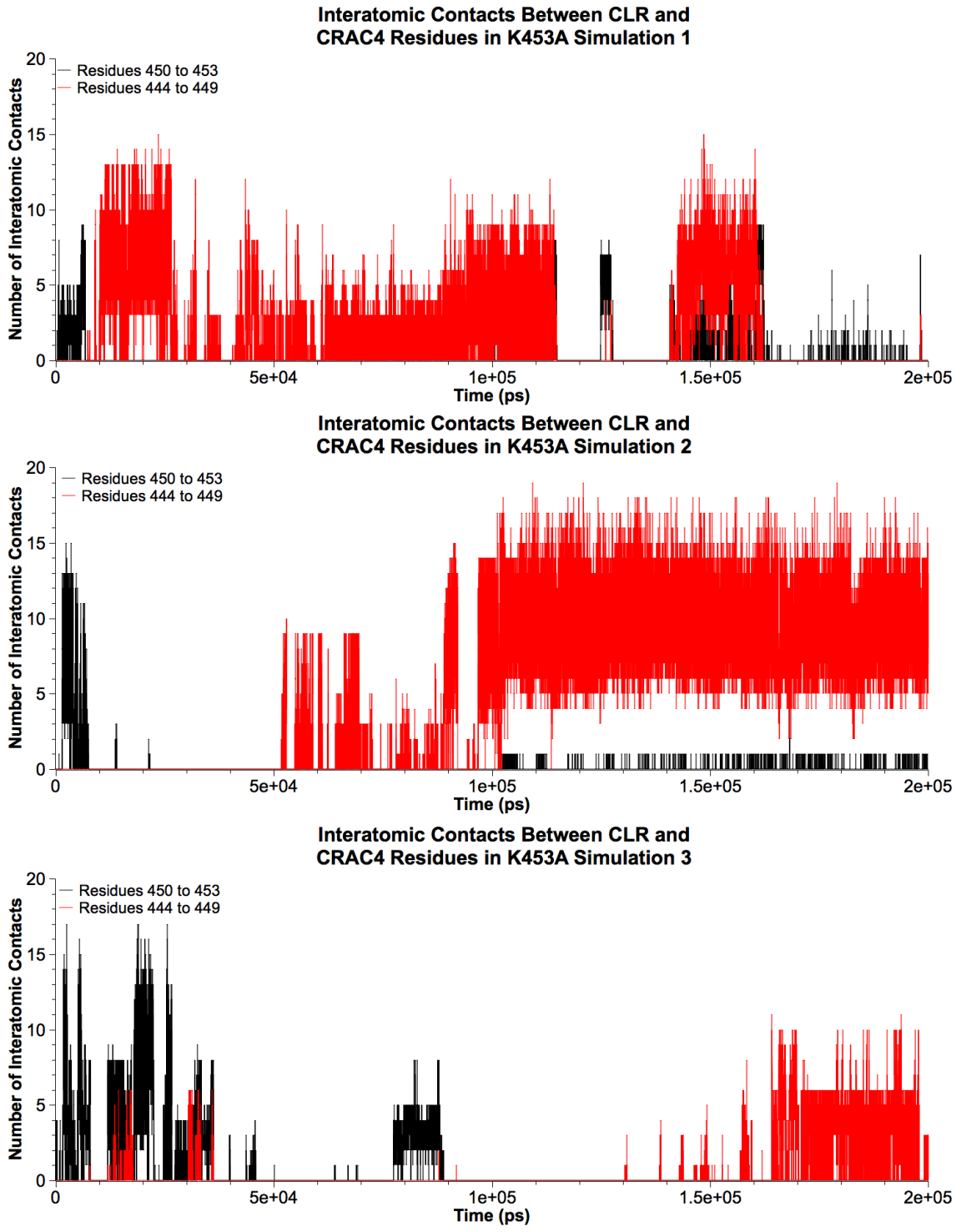


Figure 12. K453A Simulations Cholesterol and CRAC4 Interatomic Contacts Contacts less than 5 Å between CLR H9 atom and residues in CRAC4 shown as total number of contacts per frame (1ps) of each simulation.

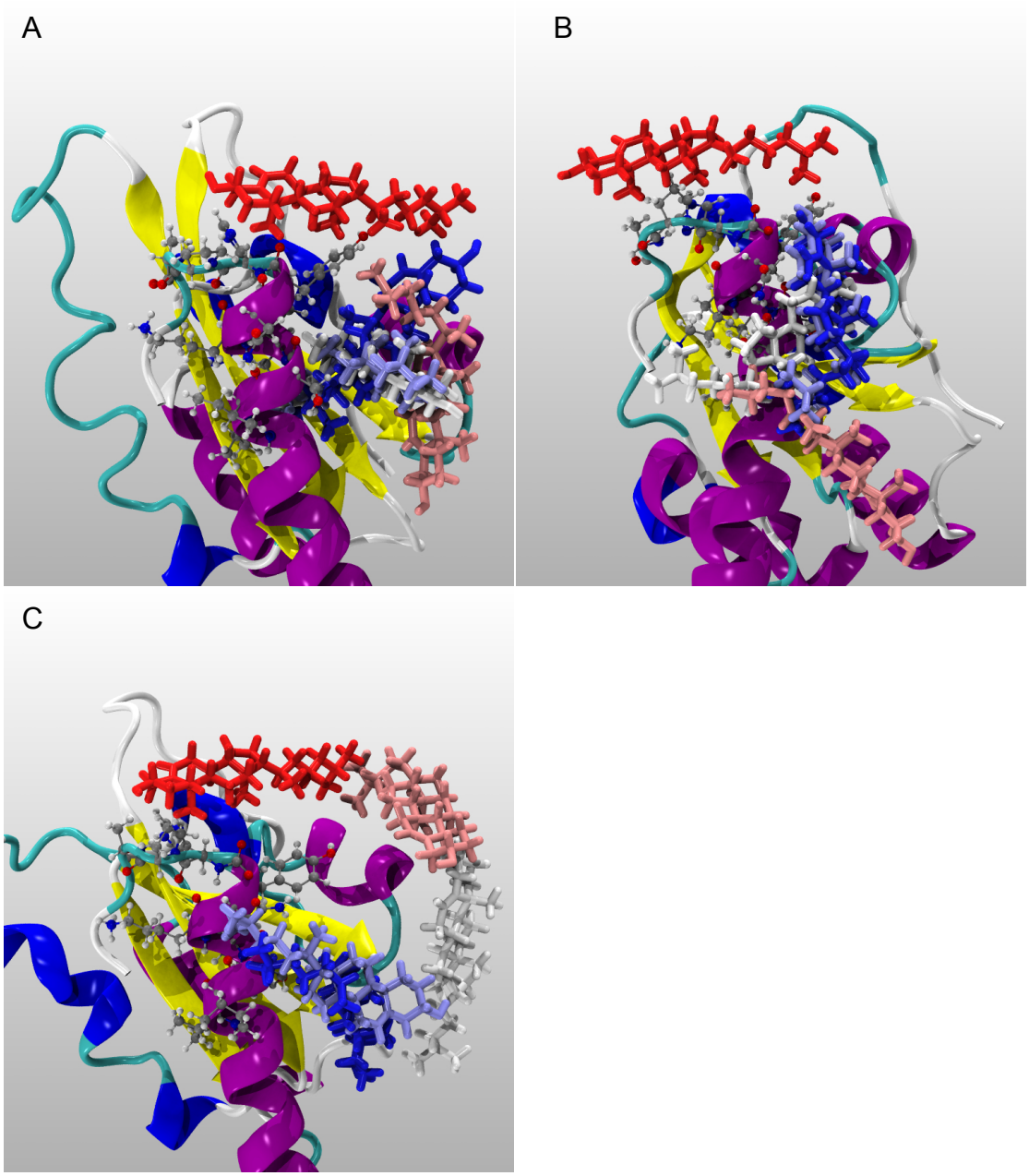


Figure 13. Positions of Cholesterol on CRAC4 Motif in K453A Simulations
CLR (stick models) coordinates in 200 ns simulations shown every 50 ns. Time steps shown are 0 ns (red), 50 ns (pink), 100 ns (white), 150 ns (purple), and 200 ns (blue). Protein is shown as ribbon structure with CRAC4 residues as ball and stick models. A) Simulation 1 has CLR move away from K453A early in simulation and remain near helical CRAC4 residues B) Simulation 2 shows a similar molecular behavior to simulation 1. C) Simulation 3 shows CLR move away from CRAC4 residues and move towards the helical residues around 150 ns.

Y450F simulations (Figure 14) showed lower occurrence of 10-15 interatomic interactions with CRAC4 residues 450 to 453 compared to trCTD simulations (Figure 10), which indicates fewer close associations with these residues during simulation time. However, the first and second simulations show sustained interatomic contacts of around 5 throughout simulation time, which corresponds to the location of CLR near Y450F shown in Figure 15 A and B. In both simulations, CLR migrates toward the mutated residue early in the simulations and in simulation 1 moves back toward K453 around 150 ns and shows a marked increase in hydrogen bonding (represented as green circles) before moving to the helical CRAC4 residues. Simulation 3 deviates from the behavior in the first and second simulations by moving quickly to the helical CRAC4 residues, but remaining far enough from the residues that 2 interatomic interactions are reported for most of the simulation with spikes of increased interaction throughout. The difference in CLR behavior in the third simulation may be explained by the difference in the energy in the system after the equilibration phase prior to MD production runs that would cause CLR to behave differently when constraints were lifted for production runs. Additionally, this could be attributed to localized pockets of low solvent density in the water box altering the small molecule's behavior. The third simulation has CLR starting in a different position and orientation than the other simulations, indicating sufficient energy was present in the system during the equilibration phases to overcome the restraints imposed (10.0 kcal/mol·Å^2) during these phases. Because of this, the distance from K453 when production runs began and restraints were lifted likely

contributed to the differential movement due to the inability of the hydroxyl of CLR to hydrogen bond to K453 in order to stabilize CLR on CRAC4 residues. Additionally, movement of CLR in Y450F simulations 1 and 2 (Figure 15) mimics the behavior of CLR in trCTD simulations 1 and 2 (Figure 11), although CLR remained in association with residues 450 to 453 longer in trCTD simulations, with CLR moving toward Y450 over time. This is distinct from K453A simulations where CLR moves toward the helical CRAC4 residues in all simulations (Figure 13).

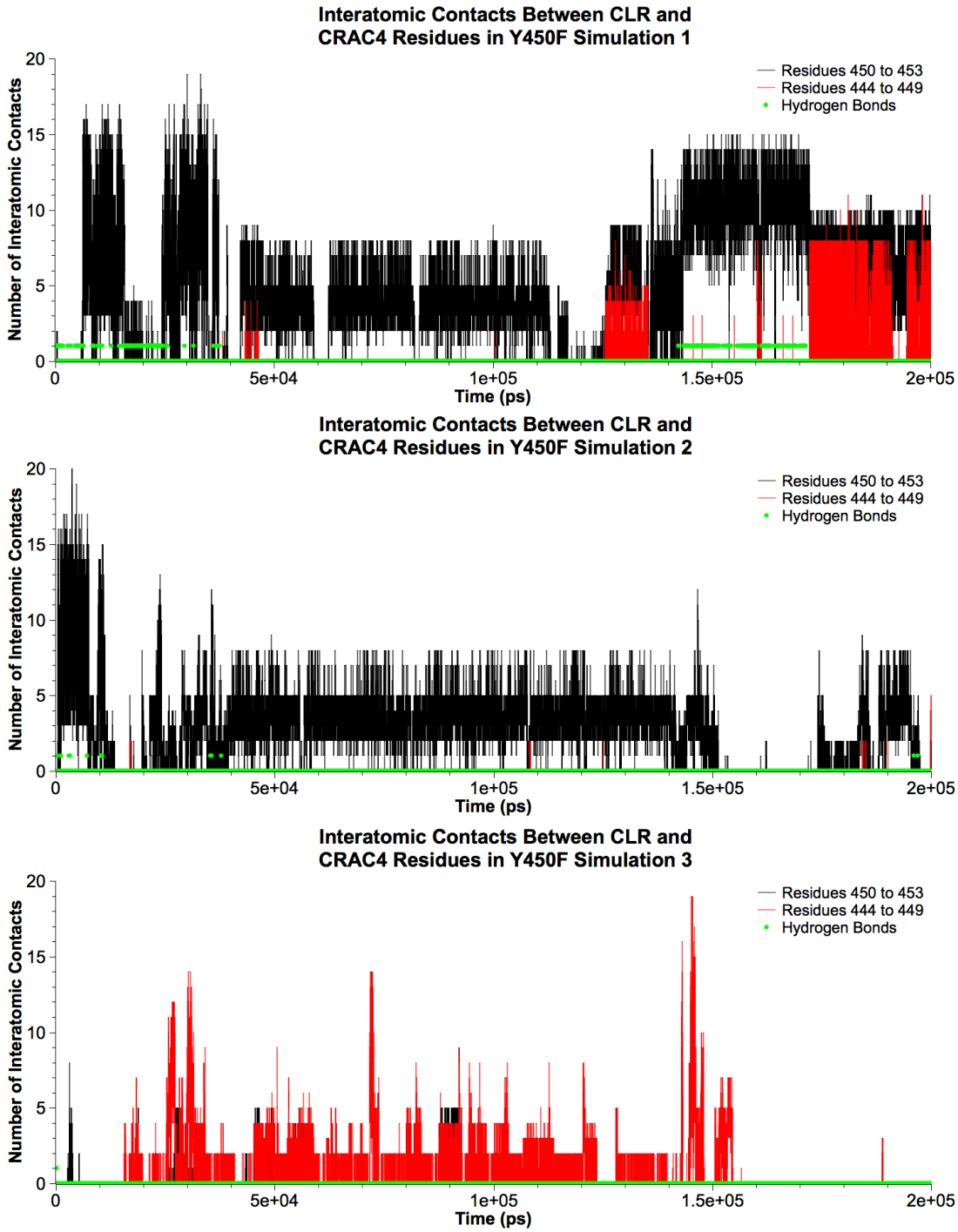


Figure 14. Y450F Simulations Cholesterol and CRAC4 Interatomic Contacts Contacts less than 5 Å between CLR H9 atom and residues in CRAC4 shown as total number of contacts per frame (1ps) of each simulation. Green circles represent points in the simulation where K453 hydrogen bonds with the hydroxyl of CLR.

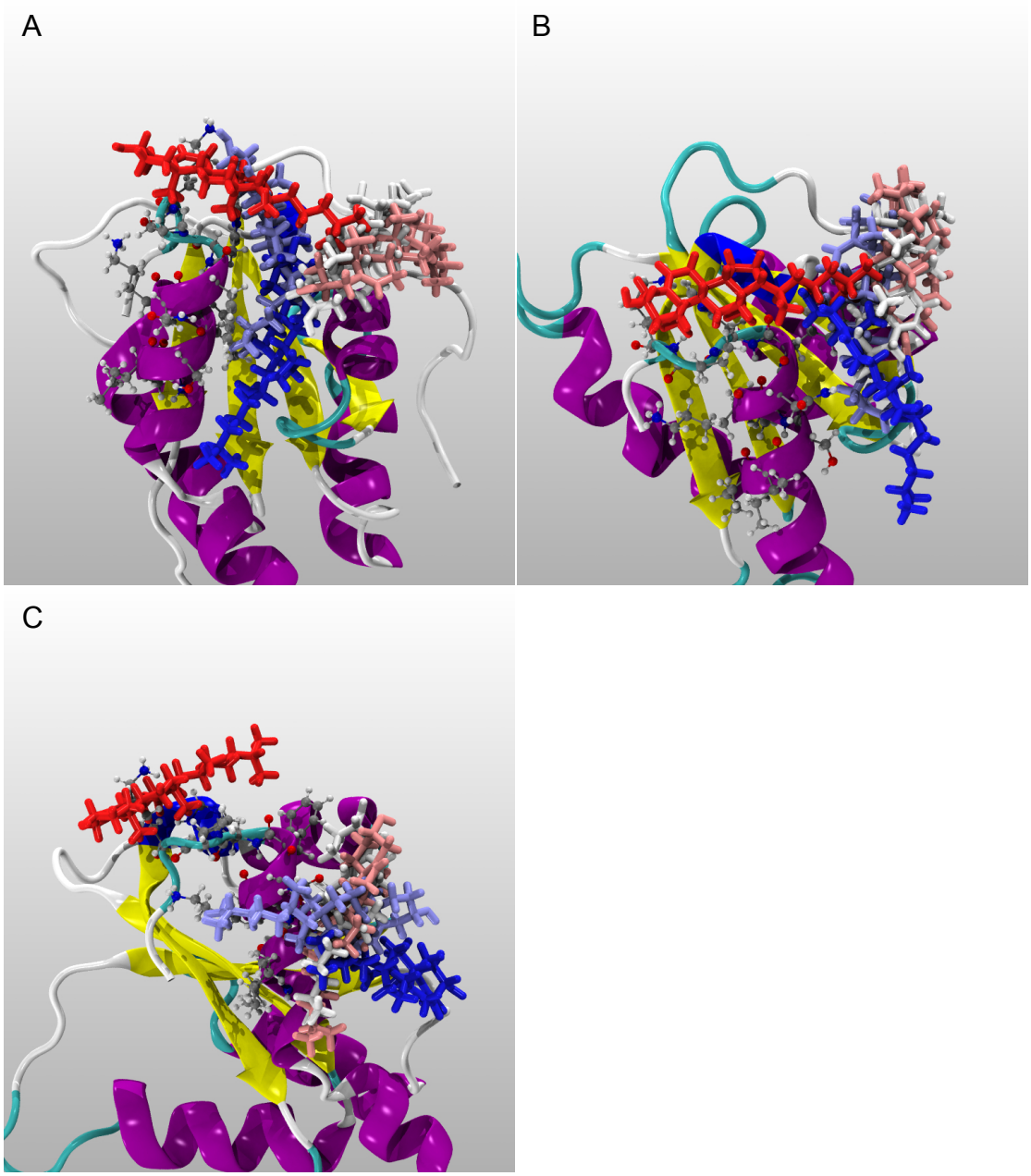


Figure 15. Positions of Cholesterol on CRAC4 Motif in Y450F Simulations
CLR (stick models) coordinates in 200 ns simulations shown every 50 ns. Time steps shown are 0 ns (red), 50 ns (pink), 100 ns (white), 150 ns (purple), and 200 ns (blue). Protein is shown as ribbon structure with CRAC4 residues as ball and stick models. A) Simulation 1 shows CLR starting on CRAC4 and migrating towards Y450F before migrating back to K453 in the last 50 ns of simulation time. B) Simulation 2 shows CLR molecule starting on CRAC4 (red) and moving away from K453, but staying near Y450F before moving towards residues 444 to 449 in the end. C) Simulation 3 shows CLR moving towards residues 444 to 449 early in the simulation.

Finally, the CTD simulations (Figure 16, Figure 17) showed CLR behavior distinct from the trCTD, K453A, and Y450F simulations. Figure 16 shows interactions between CLR and the two spatially distinct parts of CRAC4. For this set of simulations, four production runs were completed due to CLR quickly moving into the water box in the second simulation, which is indicated in Figure 16 by the lack of interactions between CLR and CRAC4 residues. This was thought to be a rare molecular event, so an additional simulation was performed. For all 4 simulations, interactions between CLR and residues 450 to 453 are lesser in these simulations compared to trCTD (Figure 10) and Y450F (Figure 14). Compared to K453A simulations (Figure 12), CTD simulations show almost no interaction with residues 444 to 449 except for a brief interaction in the third simulation with fewer than 5 interactions per frame and occurring for roughly 5 ns. In Figure 17 A and B, CLR moves to other parts of the protein distant from CRAC4 and remains there throughout simulation time. Panels C and D show CLR staying closer to CRAC4 with the third simulation; this being the most similar to trCTD and Y450F results with CLR in the fourth simulation moving toward the linker part of the protein that connects the cytosolic domain to the transmembrane segments of BK_{Ca} (Figure 1). The fourth simulation also shows CLR interacting with CRAC4 residues through the first 50 ns of simulation time before moving to the linker region.

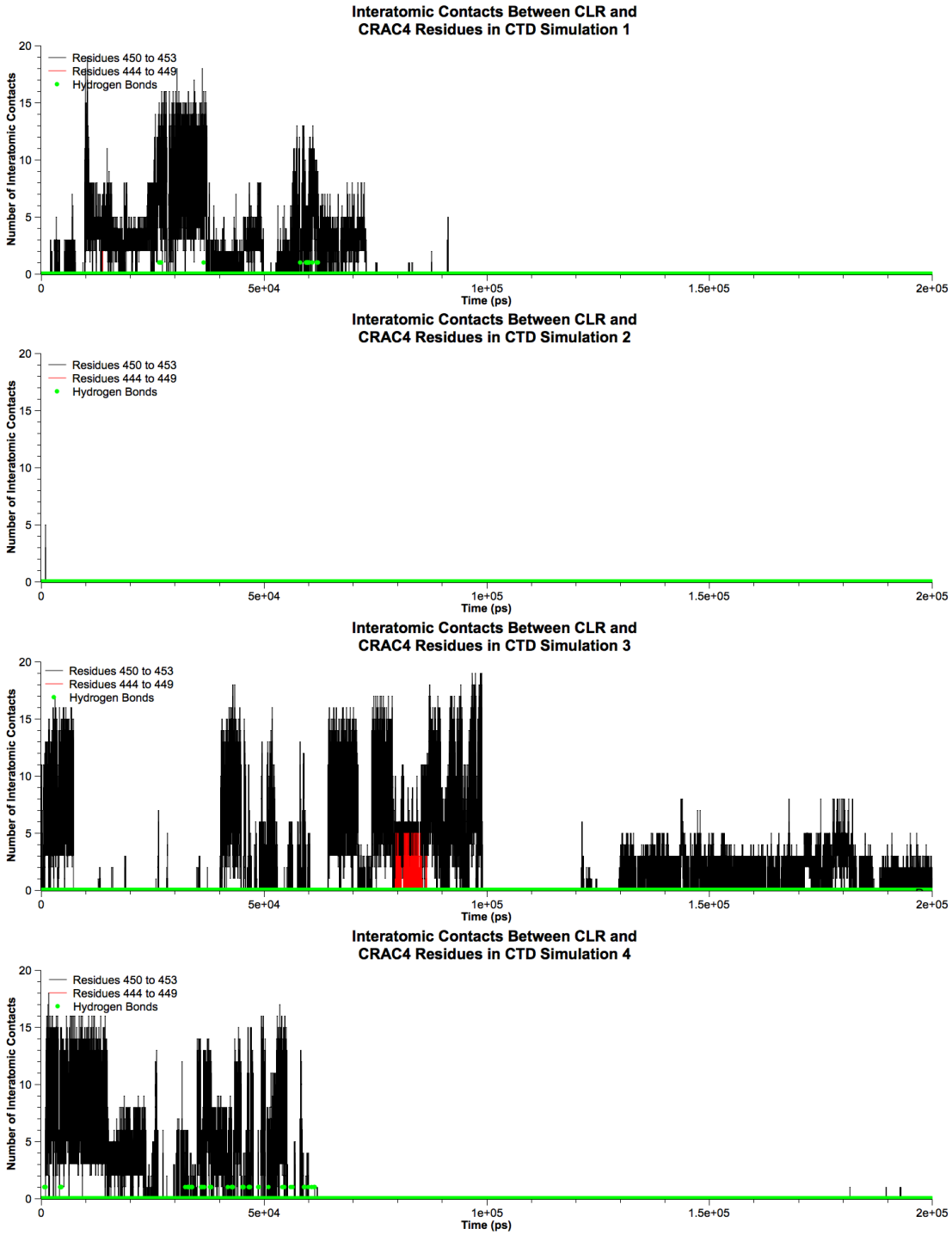


Figure 16. CTD Simulations Cholesterol and CRAC4 Interatomic Contacts Contacts less than 5 Å between CLR H9 atom and residues in CRAC4 shown as total number of contacts per frame (1ps) of each simulation. Green circles represent points in the simulation where K453 (BK_{Ca} residue K453) hydrogen bonds with the hydroxyl of CLR.

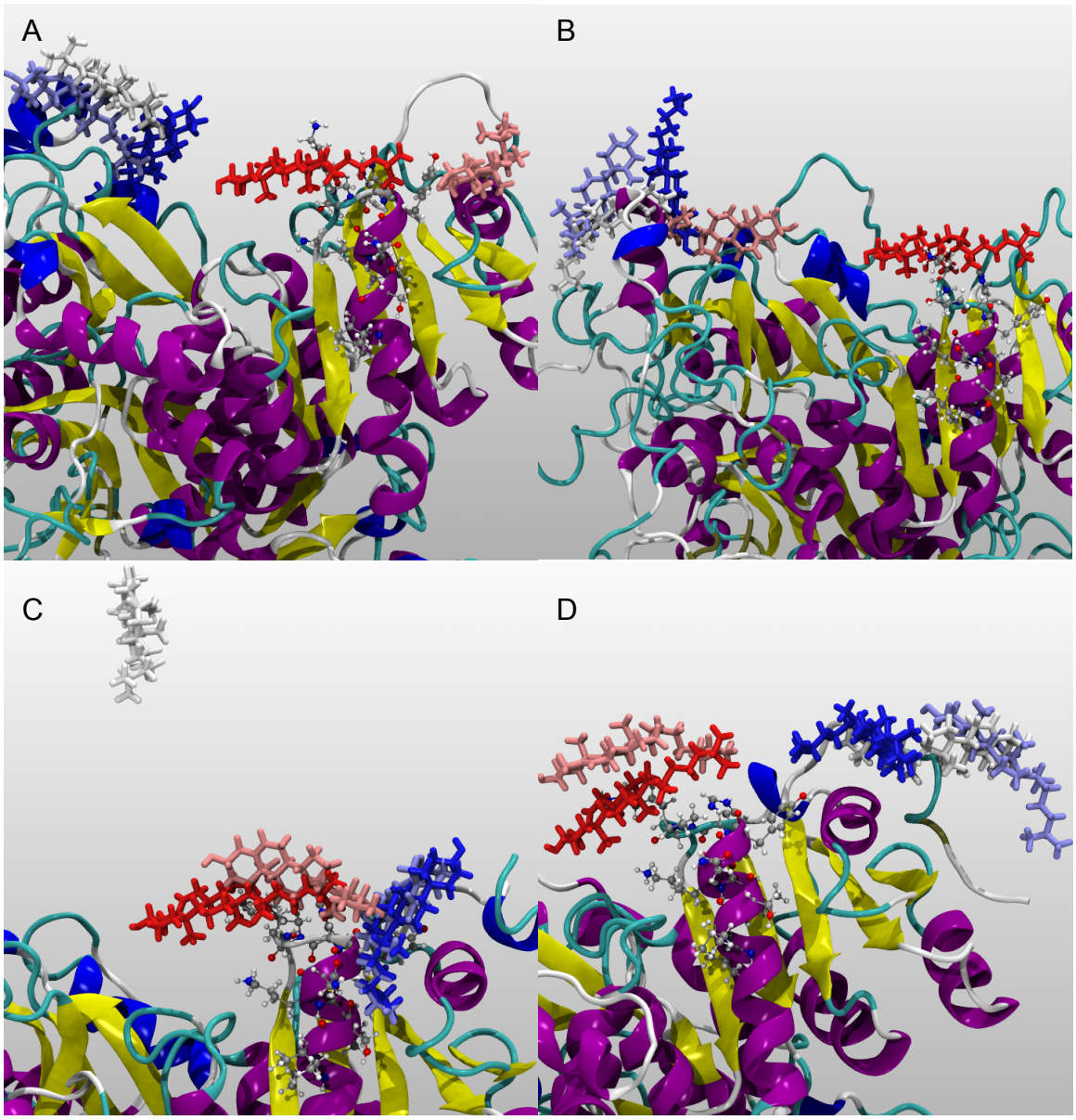


Figure 17. Positions of Cholesterol on CRAC4 Motif in CTD Simulations
CLR (stick models) coordinates in 200 ns simulations shown every 50 ns. Time steps shown are 0 ns (red), 50 ns (pink), 100 ns (white), 150 ns (purple), and 200 ns (blue). Protein is shown as ribbon structure with CRAC4 residues as ball and stick models. A) Simulation 1 shows CLR starting on CRAC4 and moving towards Y450 before migrating away from CRAC4 for the remainder of the simulation. B) Simulation 2 also shows CLR starting on CRAC4, but it quickly moves away from CRAC4 to other residues in the protein. C) The third simulation shows CLR remaining on or near the CRAC4 residues, but moving into the water box around 100 ns before returning to CRAC4. D) The fourth simulation shows CLR remaining on CRAC4 through 50 ns and then moving away to other residues in the protein.

In order to better understand CLR behavior in CTD simulations, interactions of CLR with other CRAC motifs (Table 1) were calculated for CTD simulations (Figure 18). In the first simulation, CLR moves from CRAC4 after 70 ns to CRAC7 around 100 ns exhibiting a high number of interactions, though sporadically, between CLR and CRAC7 residues. Inspection of the trajectory for this simulation shows CLR moving on and off of CRAC7 throughout the simulation, but a cause for this behavior could not be determined. The second simulation, where CLR migrated away from all CRAC4 residues very early in the simulation and moved into the water box, CLR returns to the protein and interacts with residues of CRAC6. However, CLR does not remain in contact with CRAC6 residues, moving to non-CRAC areas of the protein. There are also a small number of interactions with CRAC7 residues after 10 ns in simulation 2, but they only reach one interatomic interaction per frame and do not indicate CLR interacted with multiple CRAC7 residues.

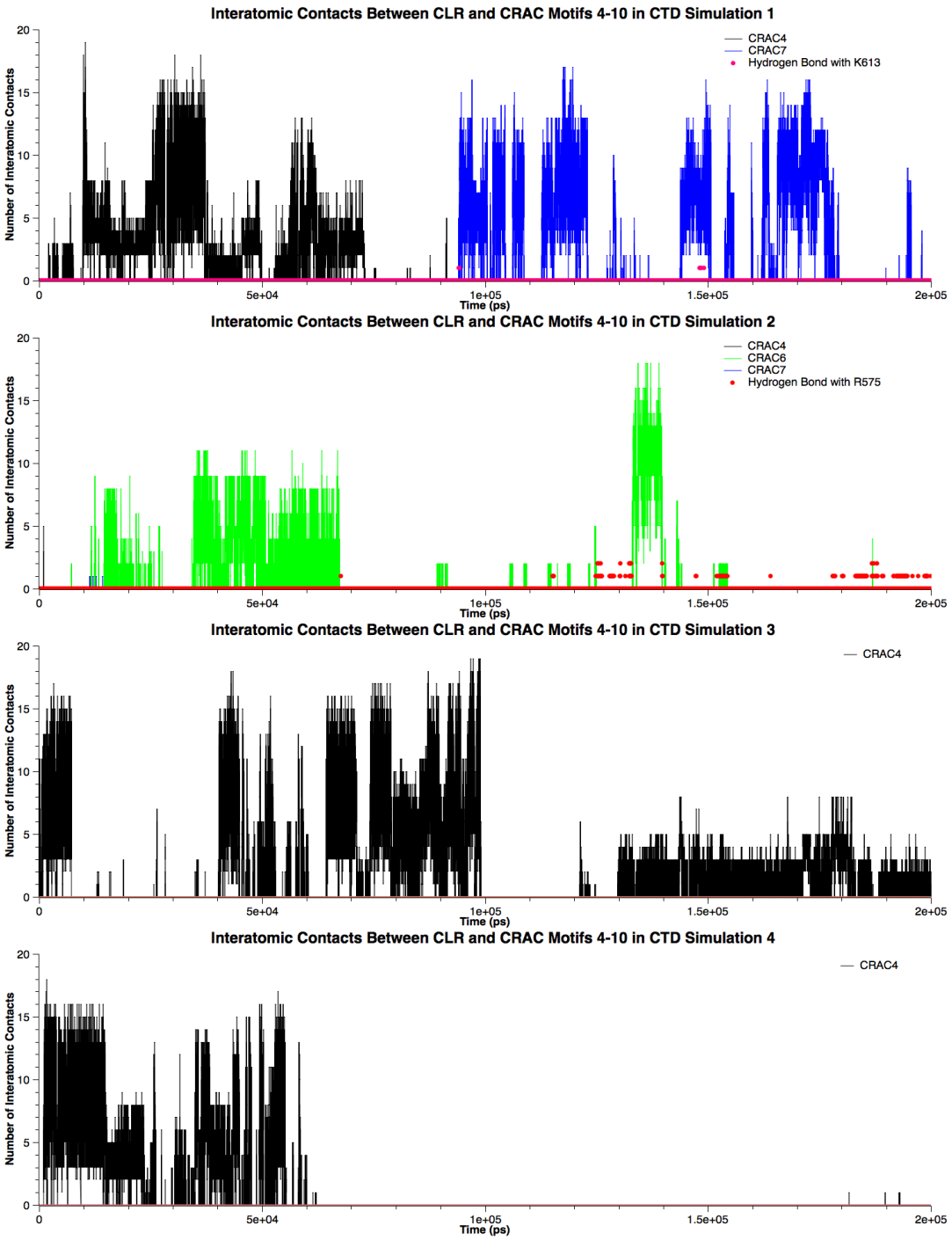


Figure 18. CTD Simulations of Cholesterol and CRAC4-10 Interatomic Contacts Contacts less than 5 Å between CLR H9 atom and residues in CRAC4 (black), CRAC5 (red), CRAC6 (green), CRAC7 (blue), CRAC8 (teal), CRAC9 (magenta), CRAC10 (orange) shown as total number of contacts per frame (1ps) of each simulation. Hydrogen bonds to R575 (red circles) and K613 (pink circles) shown for simulations where interactions with other CRAC6 or 7 motifs occurred.

2.3.3 Importance of Hydrogen Bonding for Cholesterol Interaction with CRAC Motifs

To evaluate the importance of hydrogen bonding on CLR mobility, the number of hydrogen bonds between the hydroxyl of CLR and the amine side chain of K453 was determined for all simulations except K453A, where no hydrogen bonding partner is available. The calculations were performed on the CTD, trCTD, and Y450F constructs, which all retain K453 and are capable of hydrogen bonding with the hydroxyl of CLR (Figure 8). Table 2 shows the sum of hydrogen bonds in each simulation between the hydroxyl of CLR and the amine side chain of K453 in CTD, trCTD, and Y450F simulations.

Table 2. Number of Hydrogen Bonds Between CLR and K453 in CTD, trCTD, and Y450F Simulations

	Simulation Number	Percent of Frames with Hydrogen Bonds
CTD	1	0.0315%
	2	0%
	3	0%
	4	0.218%
trCTD	1	0.833%
	2	0.205%
	3	0.388%
Y450F	1	0.0005%
	2	0.0245%
	3	0.006%

The number of hydrogen bonds in CTD simulations is considerably less than the number of hydrogen bonds in trCTD simulations. This is likely due to K453 hydrogen bonding to E479 in the CTD construct in the modeled protein

(Figure 19), which prevents the side chain from being accessible to form a hydrogen bond with CLR. The side chain interaction with E479 is an artifact of the loop modeling, which involves minimization of generated models, and structural minimization of the protein before MD simulation. The hydrogen bond between K453 and E479 is not present in the crystal structure. However, after a several nanoseconds of simulation, K453 returns to a position where it is capable of hydrogen bonding with CLR, but due to unavailability in the beginning, simulations 1, 2, and 3 have very few hydrogen bonds. Figure 16 shows that the few hydrogen bonds that occur in simulation 1 occur later in the simulation around 60 ns, while simulation 4 shows hydrogen bonding early in the simulation and also retains a high number of interactions between CLR and CRAC4 longer than any of the CTD simulations. It is probable that the lack of hydrogen bonding partner availability with residue K453 and the alternative hydrogen bonding E479 contributes to the mobility of CLR molecules in these simulations compared to the constructs based on the truncated protein.

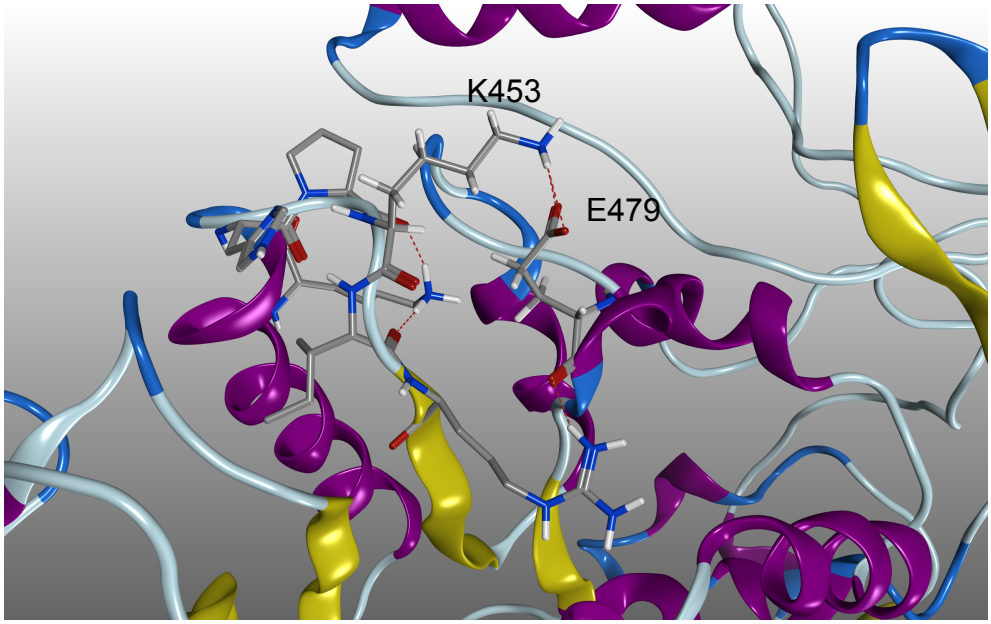


Figure 19. Hydrogen Bond Between K453 and E479 in CTD Simulation
Protein shown as cartoon ribbon structure. Residues shown as stick models with hydrogen bonds represented as dashed red lines.

In trCTD and Y450F simulations, E479 is no longer available to hydrogen bond with K453 and the residue is capable of hydrogen bonding with CLR or water molecules instead. In the two simulations with the highest number of hydrogen bonds (trCTD simulation 1 and Y450F simulation 1), the times during thesimulation where hydrogen bonds are recorded are accompanied by greater numbers of interatomic interactions between CLR and residues 450 to 453 of the CRAC4 motif; the same pattern can be seen in CTD simulations. While CLR could approach and hydrogen bond with this residues from other faces of the protein, the steady orientation, or in some cases return to starting orientation, of CLR in simulations implies CLR is drawn to these motifs. However, the underlying cause of this has yet to be determined. Combined with the data shown in Figure 12, where CLR during K453A simulations remains in contact

with residues 450 to 453 only briefly in all three simulations, these data indicate that K453 is vital to the stabilization and interaction of CLR with CRAC4 and mutation of this residue destabilizes the interaction between CLR and other CRAC4 residues in BK_{Ca} channels.

While K453 proved to be important in CLR stabilization on CRAC4, the other CRAC motifs (CRAC6 and 7) did not exhibit the same hydrogen bonding patterns as CRAC4 (Figure 18). In CTD simulation 1 where CLR migrates to CRAC7, only a small number of hydrogen bonds registered with K613, the analogous residue to K453 in CRAC7. In CTD simulation 2 where CLR migrates to CRAC6, a larger number of hydrogen bonds are recorded during the simulation with R575 and are more comparable to trCTD hydrogen bonds with K453. Around 130 ns there is an increase in the number of hydrogen bonds recorded followed by a high number of interatomic contacts, which is consistent with patterns seen in CRAC4. However, the largest group of hydrogen bonds between CLR and R575 after 180 ns do not correspond to an increase in CLR interatomic contacts with CRAC6 residues, as is seen in trCTD and Y450F simulations. This suggests CLR is approaching from one side of the protein to hydrogen bond with R575, but not approaching the other residues in the CRAC6 motif. The positions of CLR on CRAC6 will be discussed in more detail in section 2.3.5 and shown in Figure 22 These results indicate the residue analogous to K453 in CRAC6 (R575) plays a similar role to stabilize CLR on CRAC residues, but the analogous residue in CRAC7 (K613) did not show the same hydrogen bonding-interatomic interaction pattern as CRAC4 during simulation time.

2.3.4 RMSD of Cholesterol in Simulation Constructs

To further examine differences in CLR movement, the RMSD of CLR in each simulation was calculated from the starting structure of each construct for all 200 ns of simulation time (Figure 20). Both CTD and trCTD simulations show RMSDs lower than 10 Å longer than either Y450F or K453A simulations, with the exception of simulation 2 in both trCTD and CTD constructs. Y450F simulations show distinctly stable RMSDs over long periods of simulations time, which is also reflected in the snapshots in Figure 15. The high RMSD seen in CTD simulation 3 was a result of CLR colliding with the walls of the periodic box. The spikes in RMSD seen in trCTD simulation 2 and K453A simulation 2 are both caused by CLR briefly leaving the protein surface, but returning after a few nanoseconds of simulation. The K453A RMSD of around 20 for the duration of the simulations in all three simulations is consistent with the snapshots of K453A simulations in Figure 13 showing CLR staying near residues 444 to 449.

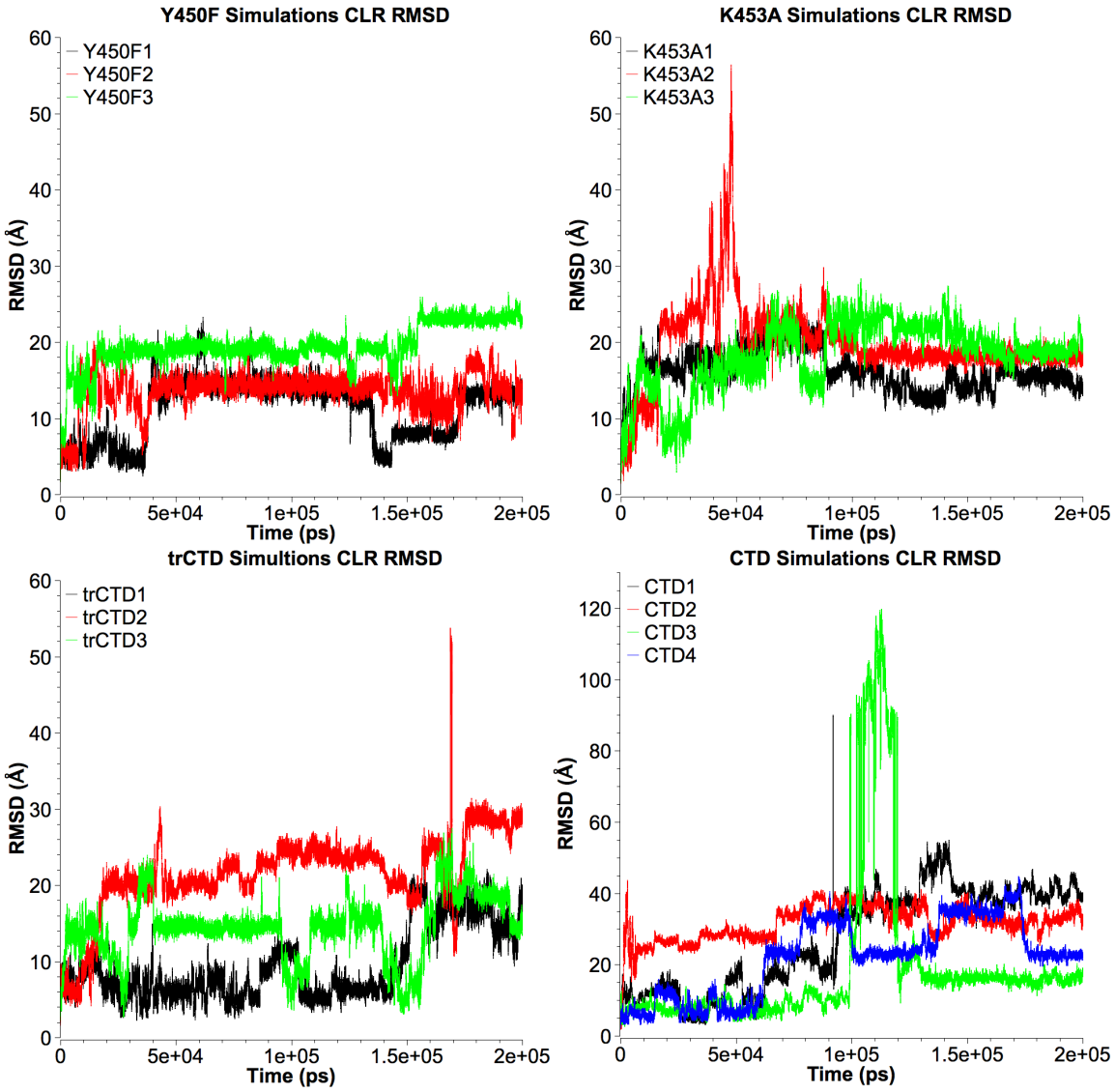


Figure 20. RMSD of Cholesterol in Y450F, K453A, trCTD, and CTD Constructs CTD and trCTD simulations show RMSDs lower than 10 Å for longer periods of simulation time than Y450F or K453A simulations.

2.3.5 Solvent Accessible Surface Area

It was observed experimentally that truncation of CRAC motifs in BK_{Ca} decreases cholesterol sensitivity[137] and CLR migrated from CRAC4 to CRAC6 or CRAC7 in CTD simulations. These two motifs (CRAC6 and CRAC7) are the CRAC motifs closest in proximity to CRAC4 that are not completely buried in the protein, like CRAC5. Hydrophobic solvent accessible surface area (SASA) was investigated as a possible explanation for the interaction of CLR with these specific sites. The CTD protein has 33.09% hydrophobic SASA, which is greater than the % hydrophobic SASA of any one CRAC motif in BK_{Ca}, with the exception of CRAC9 at 48.47%. However, because this protein has tetrameric quaternary structure and is in close proximity to the membrane by attachment through a linker to the transmembrane segments of BK_{Ca}, it is likely these numbers do not accurately represent the SASA of the assembled protein. Hydrophobic contacts between protein interfaces have previously been shown to play an instrumental role in assembly of the CTD tertiary structure,[204] as well as other globular proteins[205] and proteins that associate with cell membranes,[206] so the overall hydrophobic SASA is likely to decrease in any quaternary assembly as these areas would no longer be solvent exposed. Also, due to several loops missing from the crystal structure, the modeled sections of the protein may not accurately represent the actual tertiary structure of the protein. However, these sections did not contain any of the CRAC motifs, so the accuracy of modeled loops was not heavily scrutinized. Additionally, only one CRAC motif, CRAC4, is in close proximity to the interface between two subunits in the tetrameric assembly (Figure 21), which was determined by examining the quaternary

structure of the crystal structure, and would likely have a decreased SASA in a tetramer. However, the simulation protein file could not be assembled as a tetramer without atom clashing issues. Attempts to remedy these clashes would alter secondary structure at the subunit interfaces, so comparable calculations using quaternary structures were not possible using simulation files.

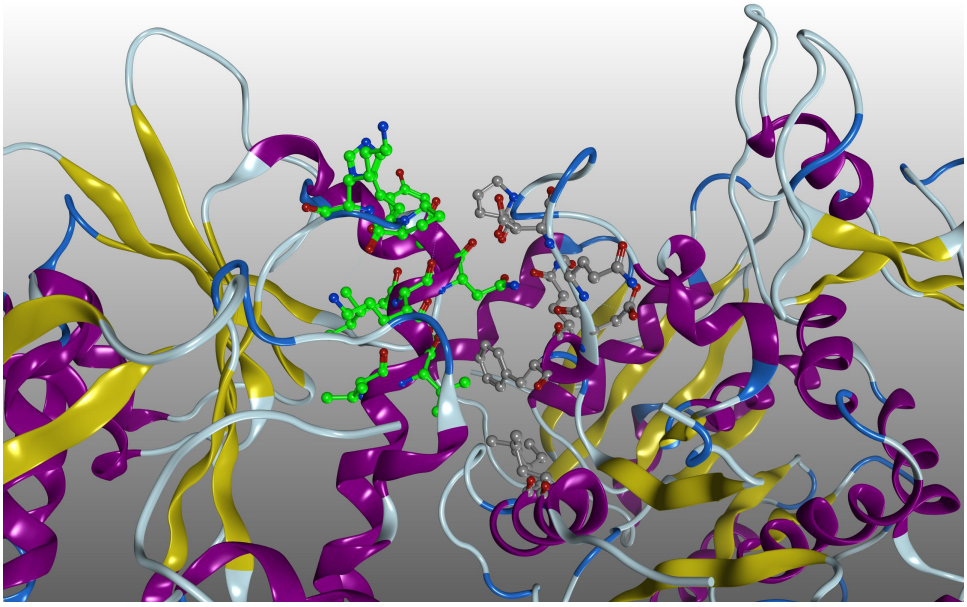


Figure 21. CRAC4 Quaternary Structure Interface
CRac4 residues are shown as green ball and stick models. Residues from the neighboring α subunit are shown as grey ball and stick models. The rest of the protein is displayed as a ribbon structure.

CRAC6 has a relatively low % hydrophobic SASA at 10.30%, which is due to the fact that most of the residues belong to a beta sheet that is buried inside the protein (Figure 22). Examination of the CTD simulation 2 trajectory shows CLR interacting with residues in parts of CRAC6 and remaining stable on CRAC6 residues before moving to another part of the protein (Figure 22). Additionally, as discussed in section 2.3.3, the time during simulation (after 180 ns) that

corresponds to the highest number of hydrogen bonds between CLR and K613 does not indicate a high number of interatomic contacts between CLR and CRAC6 residues (Figure 18). Combined with CLR positioning in Figure 22, CLR is not in contact with most CRAC6 residues during most hydrogen bonding events, but remains close enough to transiently form hydrogen bonds with K613. However, the low level of recorded contacts with CRAC6 residues is to be expected somewhat since many of the residues are buried within the protein.

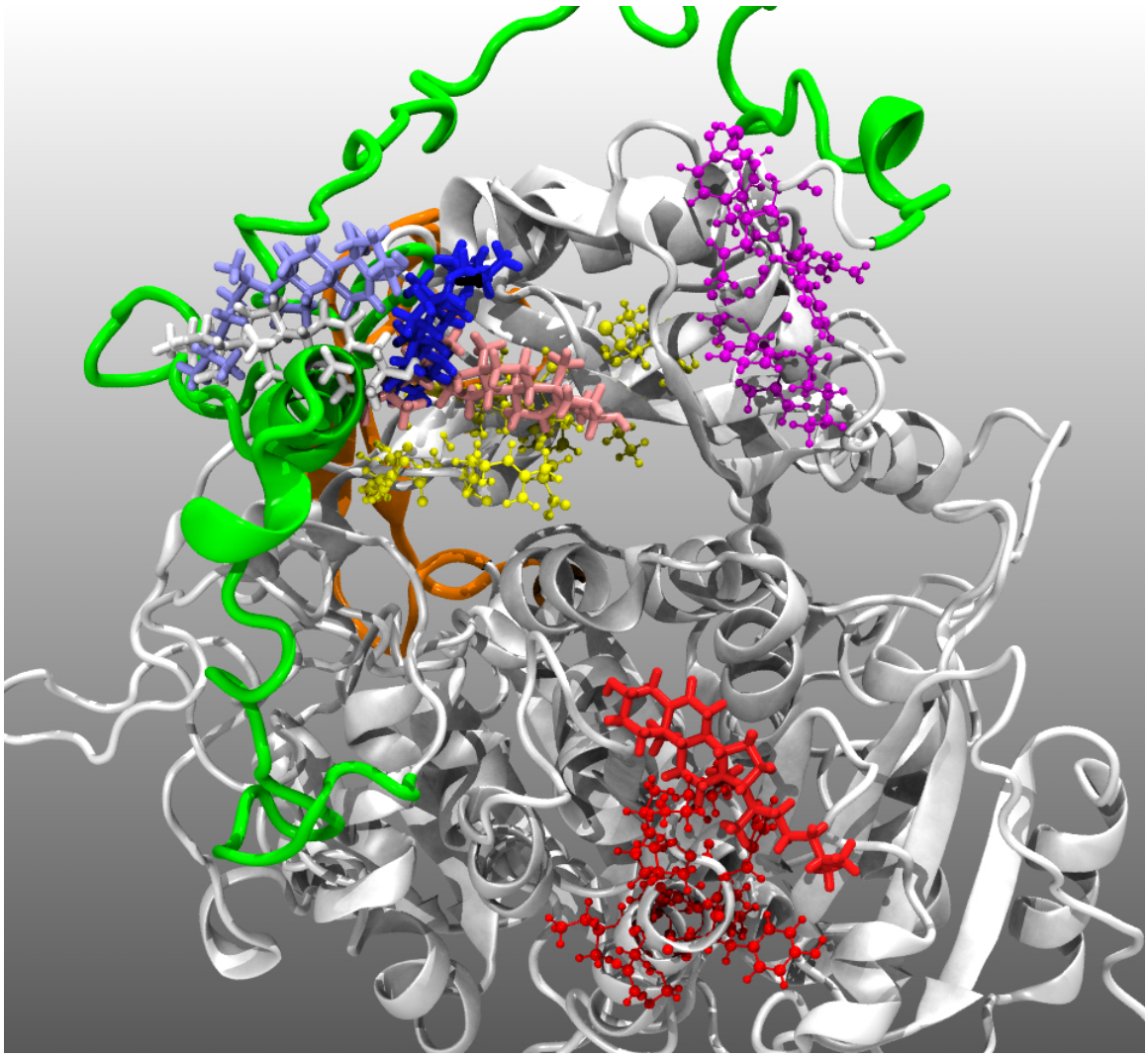


Figure 22.Top Down View of CRAC4, CRAC6, and CRAC7 Residues and Cholesterol Positions from CTD Simulation 2
All CRAC residues are represented as tubes while CLR is ball and stick representation. CRAC4 is shown in blue, CRAC6 is shown in yellow, and CRAC7 is shown in magenta. Time steps shown for CLR are 0 ns (red), 50 ns (pink), 100 ns (white), 150 ns (purple), and 200 ns (blue). Residues 287 to 387 shown in green ribbon and residues 563 to 600 shown in orange ribbon are high RMSD residues.

2.3.6 The Role of Y450 in Cholesterol Recognition

Experimentally, the Y450F mutation on CRAC4 blunted CLR sensitivity in BK_{Ca} channels[137] and comparison of simulations of trCTD (Figure 10) and Y450F (Figure 14) revealed a difference in simulation time spent in contact with K453. In Y450F simulations, CLR moved towards F450 and surrounding residues and away from K453, although it did return to hydrogen bond with K453 in Y450F simulation 1. trCTD simulations showed CLR remaining in contact with K453 through hydrogen bonding longer than in Y450F simulations before eventually moving to other parts of the protein. Despite these findings, a quantitative explanation from simulation data for the difference in CLR movement could not be determined.

2.3.7 Correlation of Molecular Dynamics Simulations and Experimental Results

Experimentally, the K453A mutant proved to be insensitive to cholesterol, which is explained in MD simulations by the lack of a hydrogen bonding partner. Without a hydrogen bonding partner present on the motif, CLR fails to remain stable on CRAC4 and quickly migrates to other areas of the protein in K453A simulations. With the role of K453 well established as a hydrogen bonding partner that serves to stabilize CLR during MD simulations of trCTD, simulation and experimental data are in agreement. The Y450F mutant was experimentally insensitive to CLR and Y450F simulations do show differential CLR movement compared to trCTD. This is in agreement with experimental data, but the simulations did not provide a possible underlying mechanism for the difference in the Y450F mutant.

Finally, the experimental observation that the truncated protein is more sensitive (about 150% compared to WT) to CLR than the WT BK_{Ca} α subunit could be explained by the absence of E479 as a hydrogen bonding partner for K453 in the truncated protein. With K453 more readily available to hydrogen bond with CLR in the truncated protein, it is possible CLR remains more stable on this protein than in the WT, as was observed in the MD simulations. While this remains a possibility, other factors could influence the difference in sensitivity between the two, including increased flexibility of the smaller protein that could bring it in closer proximity to the membrane to interact with CLR and potential decreases in CLR residence time on CRAC4 in the larger protein due to competition for interactions with other CRAC motifs which lack functional influence due to distance from the voltage sensing domain as also observed in the simulation. Truncation of the protein after CRAC4 removes the inter-subunit CTD interface shown in Figure 21 exposing residues 444 to 449, which CLR migrates to readily in many of the simulations as described in section 2.3.2. This behavior would not be possible in a fully assembled channel since these residues would be buried at the interface of the α subunits. Further studies, including simulations with explicit membranes containing CLR and a tetrameric assembly of CTD protein, are needed to determine if the difference can be attributed to K453 flexibility or other factors.

2.4 Conclusions

Simulations of the CLR sensitive BK_{Ca} α subunit cytosolic domain as WT, truncated WT, truncated K453A mutant, and truncated Y450F mutant constructs provided insight into intermolecular interactions between CLR and the protein.

While the importance of K453 was caused by hydrogen bonding with CLR that stabilizes CLR on CRAC residues, a role for Y450 and an explanation of differential CLR behavior in Y450F simulations compared to trCTD remains to be determined. Further simulations of a tetrameric assembly of CTD over longer simulation time may provide more insight into CLR interaction with CRAC residues and CLR migration between CRAC motifs.

Chapter 3

Multi-Generational Pharmacophore Modeling for Ligands to the Cholane Steroid-Recognition Site in the β_1 Modulatory Subunit of the BK_{Ca} Channel

3.1 Introduction

Functional BK_{Ca} channels result from homomeric assembly of four α (slo, slo1) subunits. In most tissues, however, the native channel complex includes modulatory β subunits that modify ion channel phenotype.[207] Due to the ubiquitous expression of the BK_{Ca} channel-forming α subunit, these channels play a significant role in many physiological functions including neuronal excitability, neurotransmission, hormonal secretion and regulation of smooth muscle tone.[2,3] BK_{Ca} α subunits consist of seven transmembrane segments[12] in addition to a long C-terminal tail that recognizes several physiologic ligands, including intracellular Ca^{2+}.[208] In contrast, BK_{Ca} β types (β_{1-4}) are small proteins, each made of intracellular N- and C-termini, and two transmembrane helices connected by a large extracellular loop.[6]

Differential expression of β subunit types across tissues allows tissue-specific BK_{Ca} current phenotypes that contribute to cell physiology in a tissue-specific manner. In particular, when co-expressed with the α subunit, the smooth muscle-abundant β_1 subunit increases the channel's apparent Ca^{2+} sensitivity and slows down channel gating kinetics.[7] In smooth muscle, an increase in the channel's apparent Ca^{2+} sensitivity allows BK_{Ca} channels to activate upon increase in intracellular Ca^{2+} levels near the BK_{Ca} channel, which are raised

during membrane depolarization and smooth muscle contraction. Thus, BK_{Ca} channel activation results in outward K^+ currents that tend to hyperpolarize the membrane and oppose depolarization-induced Ca^{2+} entry. Eventually, smooth muscle contraction is restrained and relaxation is facilitated.[209,7] Moreover, the critical role of BK_{Ca} β_1 subunit in smooth muscle physiology has been well documented: for instance, BK_{Ca} β_1 knock-out mice are characterized by increased arterial tone.[41] Decreased β_1 gene expression leads to systemic hypertension in rats[210] and increased phasic contractions in mouse urinary bladder.[211] Considering the key role of BK_{Ca} β_1 in determining smooth muscle tone, this subunit represents an attractive target for therapeutic developments against a wide clinical spectrum of prevalent human conditions associated with enhanced smooth muscle contraction, including bronchial asthma, systemic hypertension, cerebral vasospasm, erectile dysfunction, premature labor, and some bladder disorders.[131,150,212]

There are several compounds known to modify BK_{Ca} channel function, yet only a few are confirmed to act via the recently identified β_1 subunit cholane steroid-recognition site[130,141] (Figure 23). Cholane steroids such as the bile acid, lithocholic acid (LCA), increase BK_{Ca} channel activity through interaction with β_1 subunits.[130] This action is β_1 subunit-specific and not observed with $\beta_{2\text{-}4}$ subunit-containing BK_{Ca} complexes or with homomeric α channels.[133] Combined results from point mutagenesis, electrophysiology and computational modeling indicate that the cholane steroid-recognition site is located in the β_1 subunit second transmembrane domain, with T169 playing a critical role in bile acid recognition.

Previous structure-activity relationship (SAR) studies indicated that the overall shape of the bile acid, as well as the number of hydroxyls, and placement of hydroxyls and carboxylic acids, were important in bile acid interaction with β_1.[130] Among non-steroidal compounds, only one triterpenoid, 3-hydroxyolean-12-en-30-oic acid (3-HENA)[141] has been reported to specifically target β_1-containing BK_{Ca} complexes. Moreover, although there are several ligands with β_1-dependent effects on BK_{Ca} channel opening, only LCA and 3-HENA have been reported to interact with a defined recognition site[141] making the other ligands less useful in determination of a pharmacophore.

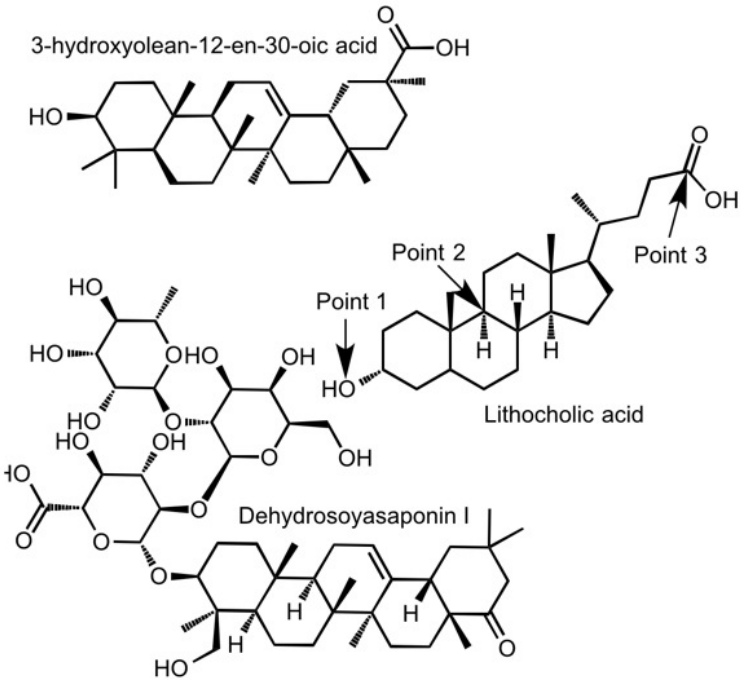


Figure 23. BK_{Ca} β_1 Selective Activators
Lithocholic acid (LCA), 3-hydroxyolean-12-en-30-oic acid (3-HENA), and dehydrosoyasaponin I (DHS-I). Points on lithocholic acid were selected for distance constraints used in initial pharmacophore searching.

High-resolution structures of the BK_{Ca} α gating-ring are available,[53] however structures of neither the complete α subunit nor any β subunit are available. Given this deficit in receptor structural data, the present study focuses on ligand-based pharmacophore model development using BK_{Ca}-ligand structures found by our groups to selectively target the cholane steroid-recognition site in BK_{Ca} β_1. As new data and techniques have become available, multiple iterations of pharmacophore modeling have led to a better understanding of how pharmacophore features effect model performance, yet lack of structural diversity amongst known ligands has posed a significant challenge. The best performing model was used to search the PubChem compound database,[213] which led to the discovery of several novel compounds that fit structural criteria posed by the β_1 steroid-sensing site while offering structural diversity over known ligands. The current study sets the stage for future *in vitro* and/or *in vivo* testing of these newly identified compounds and to further assess accuracy of the model, make further refinements, and confirm which of the candidate compounds contribute to the chemical diversity of β_1-selective agents.

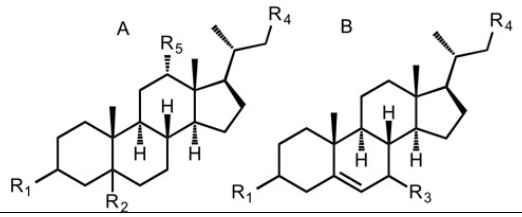
3.2 Methods

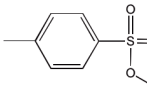
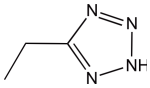
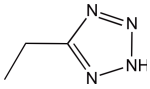
3.2.1 Pharmacophore Model Development

Initial modeling was performed using the Molecular Operating Environment (MOE)[191] 2006.08 software package. Stochastic conformational searching was performed on LCA (Figure 24 **#14**) using the MMFF94x forcefield, chosen for its ubiquitous use with small organic molecules, with gas phase solvation and a dielectric of zero. MMFF94x is an unpublished version of the MMFF94 forcefield[214] where conjugated nitrogen atoms are treated as planar

versus tetrahedral. The carboxylic acid of LCA was deprotonated to represent the expected prevalent ionization state of the molecule at physiological pH. Since LCA was reported to be the best activator among the 13 bile acids tested in a previous SAR study,[130] the lowest energy conformation of LCA was used to establish distance constraints for crude pharmacophore searching in the National Cancer Institute (NCI) database (http://cactus.nci.nih.gov/ncidb2.2/) using the 3D search feature.

Bile Acids and Derivatives



Compound # and Template	R_1	R_2	R_3	R_4	R_5
-1 B	HO	-	-	CH_2COOH	-
+2 A	HO	H	O	CH_2COOH	-
-3 A	HO	H	-	CH_2NH_2	-
-4 A	HO	H	-	CH_2CH_2OH	-
-5 A	HO	H	-		-
-6 A	HO	H	-	CH_2CH_2I	-
-7 A	HO	H	-	$CH_2CH_2SO_4H$	-
-8 A	HO	H	-	$CH_2CH_2CONH_2$	-
-9 A	HO	H	-	$CH_2CH_2CH_2NH_2$	-
+10 A	HO	H	-	CH_2CN	-
-11 A	HO	H	-		-
+12A	HO	H	-	CH_2CH_2CN	-
-13 A	HO	H	-		-
+14 A	HO	H	-	CH_2COOH	-
+15 A	HO	H	-	CH_2COOH	OH
+16 A	HO	H	-	COOH	-
-17 A	HO	H_3C	-	CH_2COOH	-
-18 A	HO	H	-	CH_2COOH	-
-19 B	HO	-	-	CH_2COOH	-
+20 A	HO_4S	H	-	CH_2COOH	-
-21 A	HO	H	-	CH_2COOCH_3	-
+22 A	HO	H	-	NO_2	-
+23 A	HO	H	-	CH_2NO_2	-
+24 A	HO	H	-	CH_2COOH	OH

Carbon Chains

Compound #	n1	R_1	n2
+25	3	$COOCH_3$	4
-26	4	$COOCH_3$	5
+27	4	COOH	5
+28	3	COOH	4
-29	2	COOH	3
-30	4	$COOCH_3$	7
-31	4	$COOCH_3$	9
-32	4	COOH	7
-33	4	COOH	9
+34	3	$COOCH_3$	5
+35	3	COOH	5
+36	2	$COOCH_3$	4
-37	2	COOH	4
-38	1	COOH	3

Miscellaneous Compounds

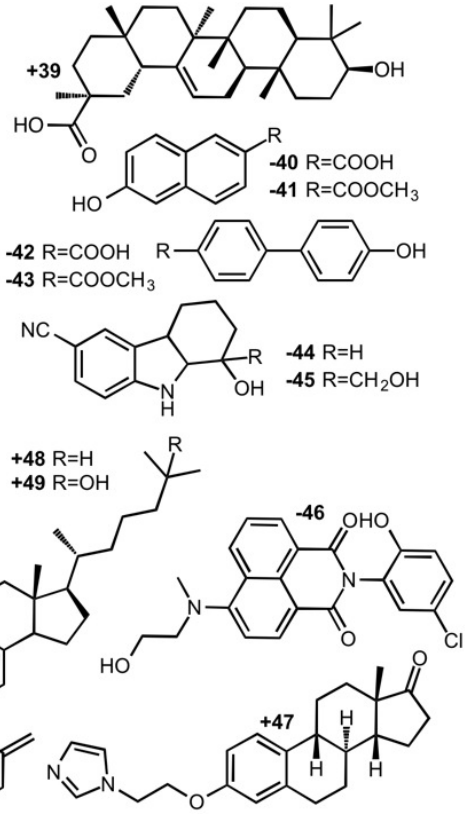


Figure 24. Compounds Used in Pharmacophore Model Development and Testing
Active compounds are indicated with a + before the compound number. Inactive compounds are indicated with a – before the compound number.

Compounds were considered “active” if they produce an increase in channel activity within 20% of the activity increase observed with LCA. Compounds were screened at either 45 μM (the EC_{50} of LCA) or 150 μM (the E_{max} of LCA).[134]

A 2nd generation model (Figure 25) was made using the Pharmacophore Elucidator feature in MOE 2009.10 to allow for searching of databases other than the NCI database. LCA and deoxycholic acid (**#24)**, previously established activators of β_1-containing BK_{Ca} channels,[134] were used as positive examples on which to train the model. Three additional bile acids that do not increase channel activity (**#17, 18, 19**) were used as negative examples on which to train the model.

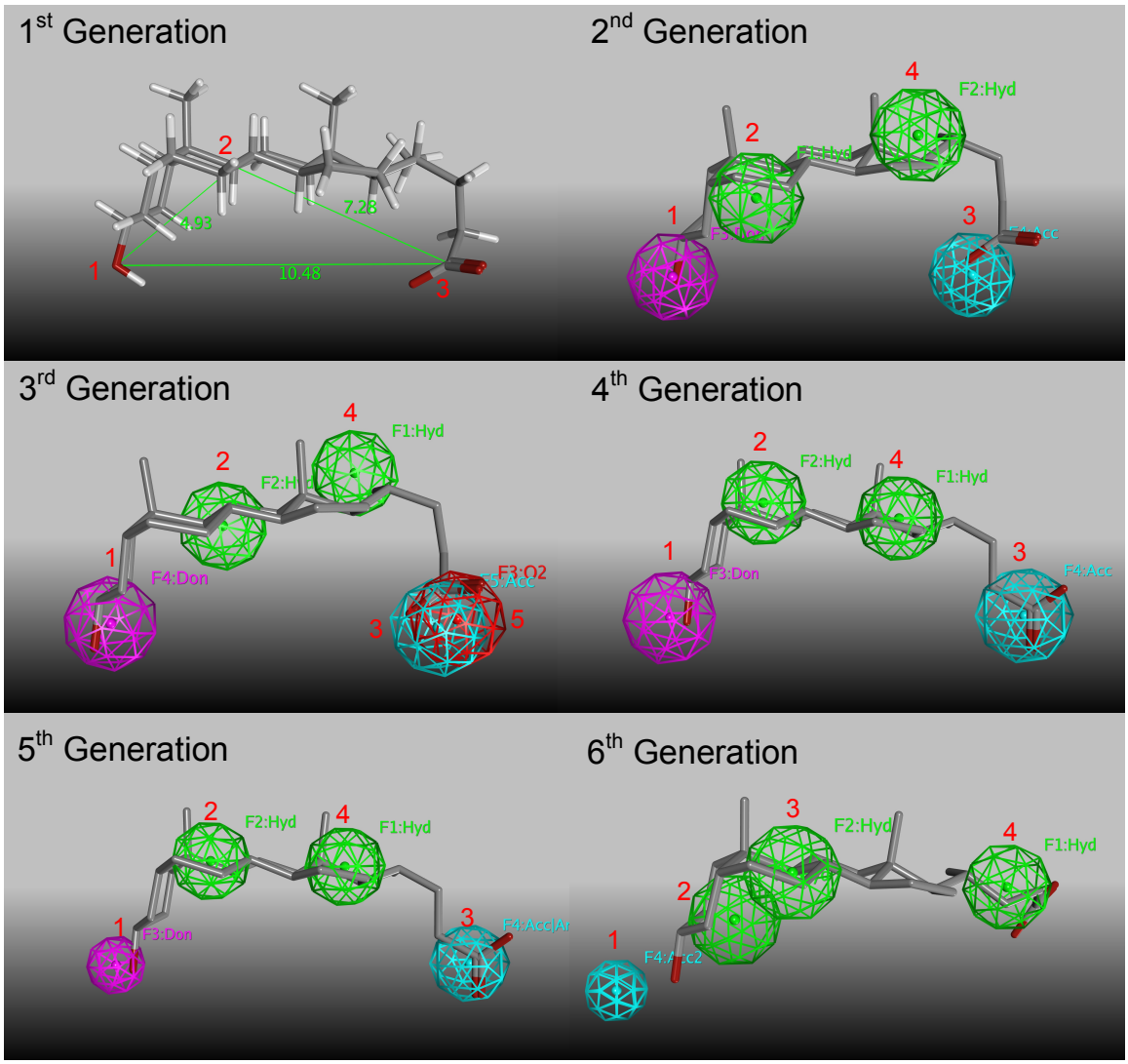


Figure 25. All Generations of Pharmacophore Models With Lithocholic Acid LCA is mapped to all generations of pharmacophore models for comparison. Magenta features represents a hydrogen bond donor, green features represent hydrophobic centers, teal features represent hydrogen bond acceptors, and the red feature represents a carboxylic acid bioisostere. In the 5th generation model the teal feature is a combination of a hydrogen bond acceptor and an anion. In the 6th generation model the teal feature is a hydrogen bond acceptor projection from overlapped hydrogen bond donors in the model training set. Numerically labeled features are described in Table 5 and Table 6. First generation crude model of pharmacophore using atomic distances between points on LCA (shown). Second generation model developed for conformational database searching. Third generation model using 5 pharmacophore features. Fourth generation model generated with more structurally diverse compounds. Fifth generation model with radius reduced pharmacophore features (see Table 5 for radii). Sixth generation model using ligand projections for a hydrogen bond acceptor (1) and three hydrophobic features (2-4).

After the discovery of 3-HENA **(#39**) as a non-steroidal β_1 activator, a 3rd generation model was developed due to the structural diversity offered by 3-HENA.[141] The 3rd generation model (Figure 25) was generated using LCA and 3-HENA as active compounds, while using two new compounds as inactive compounds to train the model (**#8, 46**) in addition to **#17**, which was used as an inactive compound in the previous generation. These compounds were selected to provide more structural diversity over inactive compounds used to train the 2nd generation model.

The 4th generation model (Figure 25) was obtained using LCA (Figure 24 **#14**), 3-HENA (**#39**), and a synthetic bile acid containing a cyano functional group in place of the lateral chain carboxylic acid (**#10**) as active compounds. Three inactive compounds were selected for structural diversity, which are shown in Figure 24 as compounds **#1, 42, 45**.

Refinements were made to the 4th generation model to yield a 5th generation model (Figure 25), as follows. 1) The radii of the pharmacophore features were reduced to be more exclusive, 2) the hydrogen bond acceptor feature was changed to include anionic compounds, 3) feature 2 was moved to align better with the steroid nucleus, and 4) features 1, 3, and 4 were made "essential" features. Thus, any screened compound would contain at least these three features.

Finally, a 6th generation model was developed using newer techniques available in MOE 2013.08 software that include ligand projection features for

hydrogen bond donors and acceptors. The same training set of compounds used to develop models 4 and 5 was used to develop model 6 (Figure 25).

3.2.2 Model Evaluation

Test set compounds and their activities were obtained from various publications[130,134,215,216] and unpublished data from our experimental collaborators, Dr. Alex Dopico and Dr. Anna Bukiya, at the University of Tennessee Health Science Center. A conformational database was generated for test set compounds using LowModeMD conformational searching[217] and the MMFF94x forcefield in MOE 2013.08. Conformations were discarded if their internal energy exceeded the lowest energy conformation by 50 kcal/mol to provide more conformational diversity over the default 7 kcal/mol cutoff. Any compounds used to train any generation of model, compounds **#1, 8, 10, 14, 17, 18, 19, 24, 39, 42, 45, and 46** were removed from the standard test set so all models would be evaluated with the same database. Thirty-seven compounds remained to calculate pharmacophore validation metrics. Sensitivity, specificity, percent yield of actives, enrichment, and accuracy were calculated for each model, as discussed in detail by Triballeau *et al*.[218] Sensitivity (Equation 1 below) represents the ability of the model to correctly pick out actives, while specificity (Equation 2) indicates the model's ability to correctly reject inactive compounds. The yield of actives (Equation 3) indicates the proportion of active molecules selected by the model. Enrichment (Equation 4) is a measure of how well the model enriches the hit list with actives compared to the proportion of actives in the whole test set. Accuracy (Equation 5) is a measure of how well the model can

correctly discriminate between experimentally active and inactive compounds. A equation for each calculation is given below:

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (1)$$

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (2)$$

$$\text{Yield of Actives} = \frac{TP}{n} \quad (3)$$

$$\text{Enrichment} = \frac{TP/n}{A/N} \quad (4)$$

$$\text{Accuracy} = \frac{TP+TN}{N} \quad (5)$$

In Equations 1-5, *TP* is the number of true positives, *FN* is the number of false negatives, *TN* is the number of true negatives, *FP* is the number of false positives, *n* is the number of compounds selected by the model, *N* is the total number of entries in the database, and *A* is the total number of active compounds in the database. In this use, a *positive* compound is one identified to

be active by the model, while a *negative* compound is identified to be inactive by the model. The term *true* indicates the result from the model agrees with the experimental result, while *false* indicates a result from the model that disagrees with the experimental result. For example, a true positive is a compound predicted by the model to be active that is experimentally active and a false positive is a compound predicted by the model to be active that is experimentally inactive. For Equations 1, 2, 3, and 5,a value closer to one is a good result. For Equation 4, a value greater than one is a good result.

3.2.3 Database Searching

Database searching was performed in a 3D compound database, which contains up to 10 conformations per molecule, downloaded from the PubChem FTP site (ftp://ftp.ncbi.nlm.nih.gov/pubchem/Compound_3D/10_conf_per_cmpd/). The compound databases were imported into MOE 2012.10 and the 3rd generation model was used to query the database. An additional search was performed using the 6th generation model to search the 3rd generation results from searching the PubChem database. The number of compounds returned from the database search was reduced by excluding anything with a LogP below 1.7 and above 5.9. These values represent the LogP (the octanol/water partition coefficient) values of compounds **#25** and **#39,** respectively, and were used because they are the highest and lowest LogP values of the most potent known activators. The molecular weights of compounds **#25** and **#39** were used to further reduce the number of compounds with a range of 198.2620 to 455.7030. Finally, since the steroid recognition site is located in the transmembrane portion

of the helix, compounds that would be too polar for membrane insertion were removed. This was accomplished by calculating the number of non-carbon heavy atoms in each molecule. Any compound with more than four non-carbon heavy atoms was removed from the hit list, since all but one known active (**#20**) contain no more than four non-carbon heavy atoms.

Clustering was performed on the remaining compounds by first calculating MACCS structural keys (a binary description of structure)[219] and similarity was determined using a 65% Tanimoto coefficient (a measure of similarity between two sets of compounds) similarity[220] threshold. The clustering overlap threshold was set at 65% or greater; meaning two compounds would require 65% or greater Tanimoto coefficient similarity of their structural neighbor lists to be placed in the same cluster. The formal charge on compounds in each cluster was used to further reduce the number of compounds by eliminating compounds with no formal charge, leaving only compounds with -1 or -2 charges to mimic the charge on the most active compounds (LCA and 3-HENA).

3.3 Results and Discussion

3.3.1 Model Selection

For the 1st generation model, three points were established on the LCA molecule (Figure 23) that were based in part on a previous study;[134] two points underscored the importance of the lateral chain carboxylic acid and the A ring hydroxyl (point 1 to point 3). A carbon in the steroid nucleus was selected as the third point to represent the concavity of LCA due to a cis A-B ring junction and is shown in Figure 25. The distances between the points and the ranges used to search the NCI database are shown in Table 3. Ranges were used to allow some flexibility outside of the values determined from a single conformation of LCA as opposed to using exact values, which would be overly restrictive. Since LCA has a LogP of 6.05, a LogP range of 4 to 8 was also used as additional search criteria to narrow the results. Due to the limitations of the browser-based search features of the NCI database, this method was abandoned in favor of more robust three-dimensional model development using MOE. While the 1st generation model provided a starting point, it was inherently limited to 3D distance descriptions of the pharmacophore and did not explicitly involve inactive compounds for model development. Results of virtual screening will be discussed in section 3.3.3.

Table 3. Distance Constraints Used in NCI Database Searching

Pharmacophore Points	Lithocholic Acid Distances	NCI Search Distances
1→2	4.93Å	4.60Å→5.80Å
1→3	10.48Å	9.00Å→11.00Å
2→3	7.28Å	6.50Å→8.30Å

The 2nd generation pharmacophore used LCA and deoxycholic acid as active compounds to develop the model and provided a more robust description of the steroid recognition site in β_1 by using four features to describe the pharmacophore in three dimensions. The set of potential models was analyzed for hydrophobic features that were spaced to convey the "bean" shaped concavity of LCA. Candidate models were also screened using the criteria that the model should contain a hydrogen bond donor feature aligned with the hydroxyl of LCA and a hydrogen bond acceptor feature aligned with the carboxylate of LCA as these two functional groups very likely play a role in the interaction with β_1. The top candidate model was evaluated for accuracy using a test set of compounds with experimentally known activity on β_1-containing BK_{Ca} channels excluding the compounds used to train all models so that all models were evaluated using the same set of compounds.

For the 3rd generation model, the top candidate model was selected from the potential models using the same selection criteria as the 2nd generation model. An additional feature, however, was allowed in the 3rd generation model:

a carboxylate ion bioisostere feature that allows other functional groups in place of the carboxylate.

The top candidate pharmacophore model in the 4th generation was selected using the previously mentioned criteria for the 2nd generation model, but the placement of hydrophobic features was more heavily scrutinized than previously. This was accomplished by ensuring any candidate models had hydrophobic features in the steroid nucleus of the active molecules and not on methyl groups.

For the 6th generation model, the two top models ranked by the model development algorithm were selected for evaluation with the test set. The models differed only in the first feature, where the top model had a hydrogen bond donor projection and the second ranked model had a hydrogen bond acceptor projection. These projections place pharmacophore features on the projected interaction sites for hydrogen bond donor and acceptor atoms in a molecule and could find a partner hydrogen bond acceptor or donator interaction partner common to all activators of the cholane steroid-recognition site. The second model was selected as the 6th generation model after evaluation of performance against the test set compounds.

Distances between pharmacophore features in models 2-6 are provided in Table 4 and the radius of each feature is given in Table 5. Because models are generated using conformational databases of actives and inactive compounds, distances between similar features are not necessarily the same. The radii of pharmacophore features were generally left at default values except in the 5th

generation model where feature radii were intentionally decreased to increase specificity.

Table 4. Distance (Å) Between Pharmacophore Features in Generations 2-6

Feature Number	2nd Generation	3rd Generation	4th Generation	5th Generation	6th Generation
1→2	4.08	4.60	5.33	5.10	5.12
1→3	9.58	9.63	11.95	11.95	7.68
1→4	8.99	9.00	8.57	8.57	13.00
1→5	-	10.18	-	-	-
2→3	7.61	7.39	9.43	9.27	2.58
2→4	5.15	4.56	4.42	4.35	8.64
2→5	-	7.76	-	-	-
3→4	5.68	5.97	5.42	5.42	7.16
3→5	-	0.63	-	-	-
4→5	-	5.96	-	-	-

Table 5. Radii (Å) of Pharmacophore Features in Generations 2-6

Feature Number	2nd Generation	3rd Generation	4th Generation	5th Generation	6th Generation
1	1.4	1.4	1.6	1.0	1.0
2	1.4	1.4	1.4	1.3	1.4
3	1.4	1.4	1.6	1.3	1.4
4	1.4	1.4	1.4	1.3	1.4
5	-	1.4	-	-	-

3.3.2 Model Performance

The metrics used to evaluate the models are shown in Table 6. These values are reported both because they represent the most commonly reported pharmacophore metrics[218] and because each gives insight into each model's performance.

Table 6. Evaluation Metrics for Pharmacophore Model Generations 2-6

	2nd Generation	3rd Generation	4th Generation	5th Generation	6th Generation
True Positives	9	5	10	9	13
False Positives	15	3	15	14	15
True Negatives	7	19	7	8	7
False Negatives	6	10	5	6	2
Sensitivity	0.60	0.33	0.67	0.60	0.87
Specificity	0.32	0.86	0.32	0.36	0.32
Yield of Actives	0.36	0.63	0.40	0.39	0.46
Enrichment	0.89	1.54	0.99	0.97	1.15
Accuracy	0.43	0.65	0.46	0.46	0.54

The 2nd generation model scored comparably to the 4th and 5th generation models (Table 6) with the standard test set in some calculated metrics, but was lower in a few metrics. Compared to the 4th generation model, this model was 0.07 lower in sensitivity, 0.04 lower in yield of actives, 0.10 lower in enrichment, and 0.03 lower in accuracy. This indicates the inclusion of more recently tested compounds with greater activity in the 4th generation training set led to improvements in model performance.

The 3rd generation model featured a fifth pharmacophore feature, which more than doubled the specificity, increased the yield of actives by 0.27, the enrichment by 0.65, and the accuracy by 0.22, but had a sensitivity almost half that of the 2nd generation model with the standard test set. Additionally, the enrichment of the hit list with active compounds was greater by 0.53 in the 3rd generation model and was the greatest enrichment of any of the models generated. The fifth feature was a carboxylate bioisostere intended to improve structural diversity in any selected compound. The low number of true positives and false positives indicate the fifth feature of this model is overly restrictive compared to four feature models. While this fifth feature did improve the model in several metrics, due to the large decrease in sensitivity, the placement and type of feature was further investigated in efforts to produce a model that better identifies the active compounds as true positives.

For the 4th generation model, the three most active compounds were used to train the model (**#10, 14, 39**). The drop in specificity, or true negative rate, of 0.54 compared to the 3rd generation model indicates that the 4th generation

model cannot correctly eliminate inactive compounds as well as the 3rd generation. This model outperformed the 3rd generation model only in sensitivity by 0.34, so further refinements were sought to improve upon the 4th generation model, without sacrificing sensitivity as was seen in the 3rd generation model.

In an effort to improve specificity of the 4th generation model, the radii of features were reduced and location of the features were moved to centralize the features in the steroid nucleus to yield the 5th generation model. While the specificity did improve to be better than the 2nd generation model by 0.04, the sensitivity, or true positive rate, declined by 0.07 compared to the 4th generation model to become equivalent to the 2nd generation model. Other metrics remained similar between 4th and 5th generation models.

The 6th generation model, which used the newer model development protocol in MOE 2013.08, provided a significant improvement in sensitivity of 0.20 over the previous best 4th generation model and 0.54 better than the 3rd generation model. The specificity of the model was on par with the 4th generation model, but 0.04 lower than the 5th generation model and 0.54 lower than the 3rd generation model. Accuracy improved to 0.54, which is 0.08 better than the 4th and 5th generation, but 0.09 lower than the 3rd generation model. Enrichment also improved over 4th and 5th generation by 0.16 and 0.18, respectively, but was 0.39 behind the 3rd generation model.

The major difference in sensitivity and specificity for the 3rd and 6th generation models demonstrates the biggest issue with pharmacophore model development for β_1 so far. With current models, any increase in sensitivity causes

a decrease in specificity. With the 6th generation being overly generous in selecting compounds as active and the 3rd generation model being much more discriminating with the test set compounds, a search with both models would eliminate a large number of inactive compounds using the 3rd generation model and then filter additional inactive compounds out using the 6th generation model (discussed in section 3.3.3 and Table 7 below). Since the PubChem database contains over 33 million compounds, the more selective nature of the 3rd generation model helps to reduce the number of compounds returned from the search of a database that large. Subsequent screening with the 6th generation model serves to filter additional inactive compounds while retaining the majority of active compounds due to the very high sensitivity of the model.

3.3.3 Virtual Database Screening Results

The 1st generation pharmacophore search of the NCI database yielded 6 compounds, but due to limitations on availability from NCI none of the compounds were available for screening. Also, the limitations of the web browser-based query in the NCI database with distances assigned to atom types led us to pursue a 3D pharmacophore model with feature annotations. However, the distance constraints established for the NCI search were later used to guide development of potential 2nd generation models. Development of 3D models also allows for searching of databases other than NCI.

A summary of the 3rd generation search and filtering results along with the results from the 3rd and 6th generation combined search are provided in Table 7. Since the 3rd generation model had the highest specificity and enrichment, and the PubChem database contains more than 33 million compounds, this model

was selected to reduce the size of the potential hit list from pharmacophore searching. Additionally, since compounds must be assayed using the patch clamp technique, which is time-consuming and not a high-throughput screening method, fewer compounds are preferred to reduce the labor and cost associated with this experimental technique. Initially, there were 317,771 compounds returned by the model. Because the laborious nature of visually inspecting over 300,000 compounds, filters were applied to the hit list to further reduce the number of compounds returned by the pharmacophore search. The upper and lower limits of the best known activators were used to filter by molecular weight, LogP, and the number of non-carbon heavy atoms. This removed any compounds that were unlikely to be activators based on our current understanding of the pharmacophore by removing compounds that would be too large for the recognition site or too polar for membrane partitioning. Since the inactive compounds have a larger LogP range of 0.47 to 8.0 and a larger molecular weight range of 142.154 to 516.787 compared to a LogP range of 1.7 to 5.9 and a molecular weight range of 192.262 to 455.703 for active compounds, using these filters should further remove inactive compounds. Additionally, restricting the number of non-carbon heavy atoms reduced the number compounds that would not insert well into the membrane to access the steroid recognition site. The total number of compounds was reduced to 7,582 after all filters were applied.

Table 7. Number of Compounds Returned by Pharmacophore Searching

	3rd Generation	3rd Generation and 6th Generation	Percent Difference
Unfiltered Compounds	317,771	243,119	30.7%
LogP and Molecular Weight Filtered	52,248	41,148	27.0%
Non-Carbon Heavy Atom Filtering	7,582	6,962	8.9%
Number of Clusters	953	779	22.3%

Using the combined 3rd generation and 6th generation models to search produced 30.7% fewer compounds than the 3rd generation model alone before any filters were applied. After filtering by LogP, molecular weight, and compounds with fewer than four non-carbon heavy atoms the difference in the two databases decreased to 8.9%. Since the percent difference is not that large after all filters were applied, the combined search approach did not provide a significant advantage over just using the 3rd generation model and filters for database searching, so screening compounds were selected from the 3rd generation model search results without secondary filtering by the 6th generation model. If a smaller database were being utilized for searching that did not require additional filters to produce a manageable number of results, the two model approach would provide the advantage of reducing the number of compounds by ~30%, with most of the compounds being inactive due to the high sensitivity of the 6th generation model.

Structural similarity clustering of the 3rd generation model search results was accomplished using MACCS structural keys to fingerprint molecules and Tanimoto similarity. A total of 953 clusters were produced at a 65% similarity threshold with the largest cluster containing 5,443 of 7,582 compounds. The formal charge calculations provided 7,193 compounds with a charge of -1 or -2. These cutoffs were selected due to the -1 charge on the two most potent compounds, LCA and 3-HENA. However, because the synthetic LCA derivative cyano compounds are active, this method had the potential to remove compounds that could be activators of BK_{Ca} through β_1.

All clusters were visually inspected, but the largest cluster (cluster 1) contained bile acid derivatives and triterpenoids similar to LCA and 3-HENA, as well as many compounds with structural diversity over current test and training set compounds (Figure 24). During visual inspection, bile acids, carbon chains similar to compounds **#25-38,** compounds similar to 3-HENA, or compounds with more than two hydroxyl groups in the portion of the molecule that should be inserted into the membrane were removed as possible candidates in an effort to improve the structural diversity of tested compounds. Since the structure-activity relationships of bile acids have been extensively characterized already,[130,134] further testing of additional related examples would not prove as useful as more structurally diverse compounds (Figure 26). The choice of two hydroxyls was made to maintain low enough polarity for membrane insertion since a previous SAR study indicated that more than one hydroxyl facing the β_1 transmembrane region can reduce a ligand's biological activity, yet removing all hydroxyls from

the bile acid molecule abolished bile acid activation of β_1 subunit-containing BK_{Ca} channels.[130] Optimally, a compound would have no more than one hydroxyl, but the single hydroxyl cutoff proved too restrictive.

Thirty-three compounds were selected by visual inspection, but since PubChem database contains both commercially and non-commercially available compounds, only 16 (Figure 26) had commercial vendors and will be purchased for *in vitro* testing. Although many compounds selected by visual inspection were not commercially available, the seven compounds denoted with an asterisk on the right side of Figure 26 are commercially available alternative compounds that are structurally similar to ones that were not available. These compounds were suggested by PubChem as structurally similar to the originally selected compound and were also in the compound hit list. Since many compounds are structurally similar in PubChem, only one compound was picked from groups of structurally similar compounds, but commercial availability was not determined until after visual inspection of the hit list.

Ongoing *In vitro* testing by our experimental collaborators of compounds selected by the 3rd generation pharmacophore will provide final validation of the value of the model in selecting active compounds and will help to determine whether these candidate compounds are indeed β_1-selective agents. Any new compounds with novel substructures will fuel further improvements in this pharmacophore model or future generations of models to produce a robust virtual screening method to find β_1-selective agents.

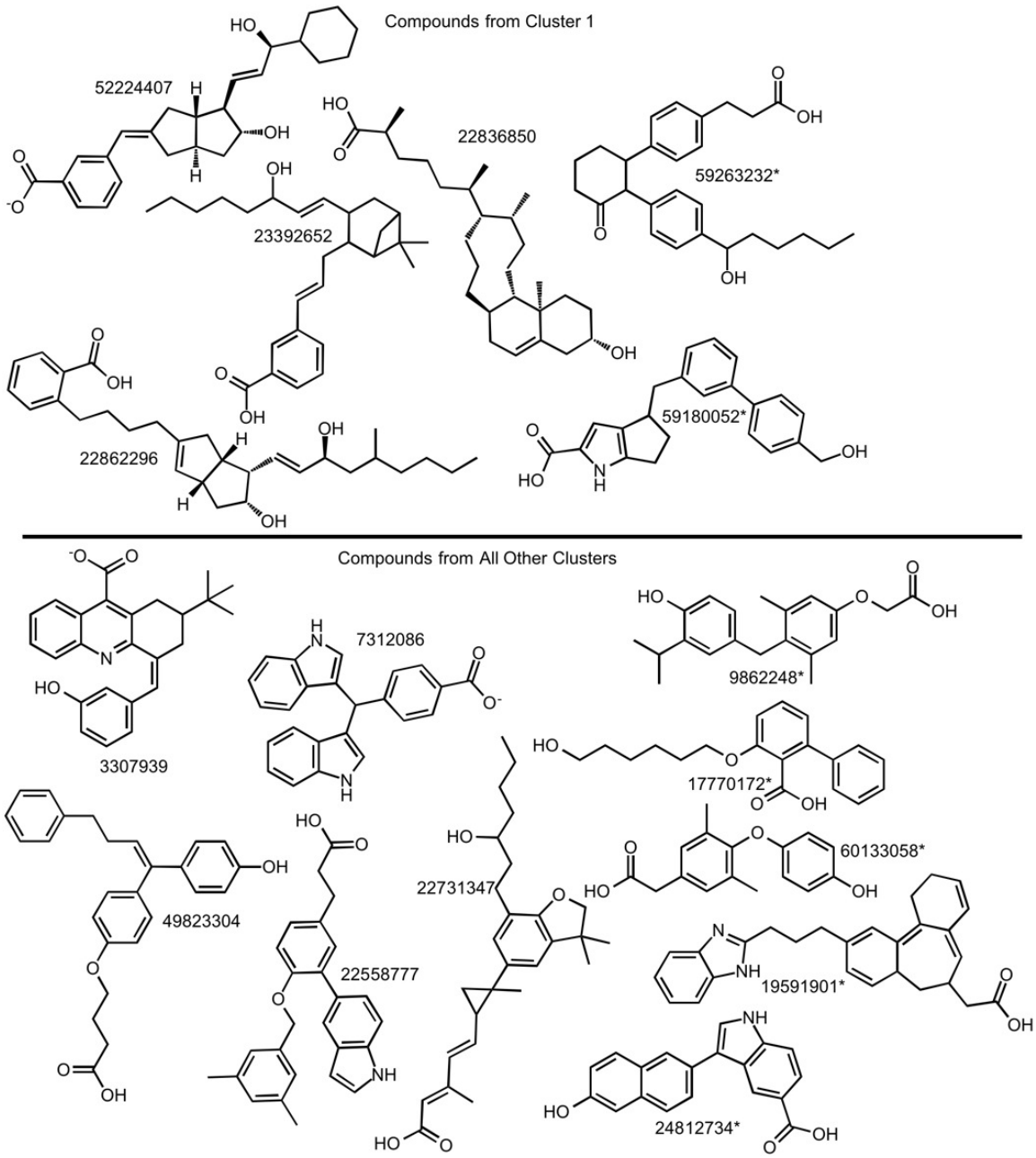


Figure 26. Commercially Available Compounds Selected from PubChem Database with Compound ID (CID) Numbers

Compounds were selected by visual inspection of structural similarity clusters. Compounds above the black line came from the largest cluster of 5,443 compounds. Compounds with an asterisk (right side of figure) are commercially available alternatives suggested by PubChem for compounds that were not available. These compounds were verified to have been selected by the model and present after all filters were applied.

3.4 Conclusions

Our pharmacophore model represents a first step in developing a robust tool for the discovery of novel activators of the β_1 steroid recognition site. While iterative improvements in pharmacophore modeling and new information have advanced the pharmacophore model over time, further improvements are still needed to address the issues with specificity and enrichment without the need for additional filtering.

Chapter 4

Expression, Purification, and Functional Studies of the Modulatory β_1 Subunit of the BK_{Ca} Channel in *E. Coli*

4.1 Introduction

With recent studies showing that genetic mutations[221,84] or genetic knockout[41] of the β_1 subunit of the BK_{Ca} channel have a deleterious effect on the subunit's physiological function, understanding the protein's structure will lead to new therapeutic options for treating BK_{Ca} channel-linked diseases. Thanks to recent advancements in stabilization of purified membrane protein samples and solution nuclear magnetic resonance spectroscopy (NMR) data acquisition,[222] structural characterization of the β_1 subunit has become more feasible. However, the protein must first be expressed, purified, and shown to possess native tertiary structure through functional studies. While the most relevant β_1 proteins to structurally characterize are derived from mammalian expression systems, because of the low yield and high cost of mammalian protein expression systems, this approach is not viable. Additionally, the large quantity of protein required for solution NMR experiments makes protein expression in *Escherichia coli* is the most viable, cost-effective option.[223]

Since *E. coli* utilize different and fewer post-translation modifications, which are used for protein structure and trafficking in mammalian systems,[224] the *E. coli* expression system can sometimes create unusable protein samples with incorrect folding or insoluble aggregates called inclusion bodies. However, *E. coli* bacterial expression systems have been used extensively to successfully

produce protein samples for structural characterization by X-ray crystallography, NMR, or cryo-electron microscopy with mammalian expression systems accounting for less than 0.5% of proteins deposited in the Protein Data Bank.[223]

Eukaryotic membrane protein expression in *E. coli* has faced significant challenges ascribed to disruption of cellular processes sometimes leading to cell death or halted cell growth and division.[180] However, newer methods to control *E. coli* stress responses to eukaryotic membrane protein expression[181] and the development of new expression vectors using various native *E. coli* proteins as fusion protein partners[225] have led to improvements in yield. Over-expression of membrane proteins as inclusion bodies, which typically requires denaturation and lengthy refolding procedures, is traditionally viewed as sub-optimal because of low yields of functional protein.[226] However, newer methods[227,228] allow for more efficient, higher yield production of proteins from inclusion bodies.

While traditional expression vectors, like pET-28a (Novagen), contain useful protein tags for purification, newer expression vectors that offer enhanced expression through protein fusion partners serve to improve expression and purification of many recombinant proteins from *E. coli* cells.[225] For example, the use of maltose binding protein (MBP) enhances the solubility of many proteins while also serving as an affinity chromatography tag that can bind to dextrin sepharose beads to aid in purification from cell lysate.[229] While the hexahistidine tag (6x-His tag) has seen ubiquitous use for immobilized metal affinity chromatography (IMAC) purification,[229] protein samples are not typically >95% pure after IMAC, so these solubility tags are typically used in conjunction with

other common forms of purification including size-exclusion and cation exchange chromatography.

The vectors utilized in this study to express protein are shown in Figure 27. The pMCSG series of vectors were developed by the Midwest Center for Structural Genomics and the two used in this study, pMCSG19 and pMCSG29, both contain MBP fusion tags, a tobacco vein mottling virus (TVMV) protease cleavage site, a His tag for IMAC purification, a tobacco etch virus (TEV) protease cleavage site, and an ampicillin selection marker.[225] The inclusion of the two different cleavage sites allows for selective cleavage of only the MBP fusion protein or both the His tag and the fusion partner. The vectors vary in that pMCSG19 has the MBP fusion on the N-terminal end of β_1 and contains a 6x-His tag, while pMCSG29 has the MBP fusion on the C-terminal end of β_1 and has a 10x-His tag. Both of these vectors utilizate ligation independent cloning (LIC) techniques to insert the desired gene by utilizing long DNA overhangs to bind the gene and plasmid together, which negates the need for the long DNA ligase step of traditional cloning.[230,231] Additionally, pE-SUMOpro (LifeSensors) utilizes a small ubiquitin-like modifier (SUMO) fusion tag that enhances solubility and expression of some proteins by adding a native E. coli chaperone protein to the desired protein.[232]

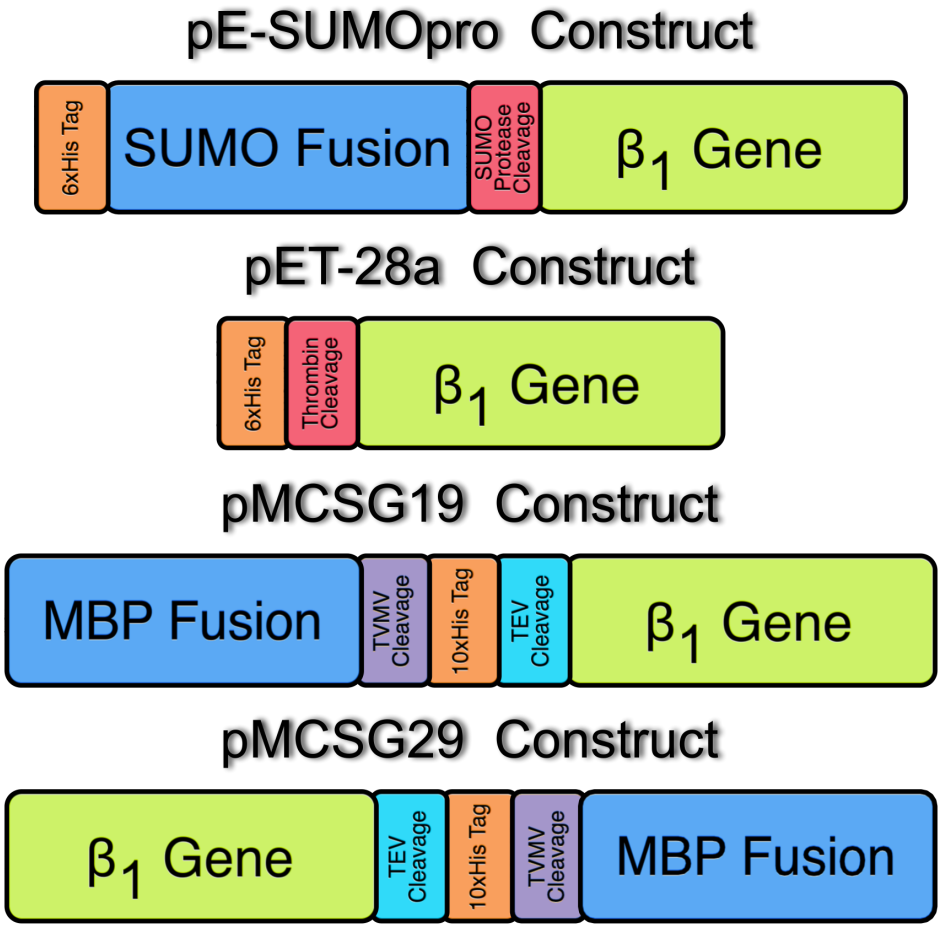


Figure 27. Bacterial Protein Expression Vector Constructs
These vectors were the first two constructs used for expression trials of the β_1 subunit of the BK_{Ca} channel. pE-SUMOpro contains a 6x-His tag (orange) for affinity chromatography purification, a SUMO fusion protein (blue), a SUMO Protease 2 cleavage site (red) to remove the fusion partner, and the inserted gene (green). The pET-28a construct contains a 6x-His tag (orange), a Thrombin protease cleavage site (red) to remove the 6x-His tag, and the inserted gene (green). The pMCSG19 construct contains an N-terminal maltose binding protein fusion partner (blue) followed by a TVMV cleavage site (purple), a 6x-His tag (orange), a TEV cleavage site (teal), and the β_1 gene (green). pMCSG29 contains the β_1 gene (green), a TEV cleavage site (teal), a 10x-His tag (orange), a TVMV cleavage site (purple), and a C-terminal maltose binding protein fusion partner (blue).

To further optimize expression of membrane proteins, *E. coli* cell lines BL21 (DE3) and C43 (DE3)[233] can be utilized. BL21 cells are standard *E. coli* expression cells, while the C43 line is a derivative of BL21 that was shown to overexpress membrane proteins better than BL21.[233,234] While C43 cells were first published in 1996, it was recently discovered that C43 cells contain three mutations in T7 RNA polymerase that delay the onset of transcription therefore slowing the production of protein and potentially reduce formation of insoluble protein aggregates.[182] These cell lines, along with the four expression plasmids, were used to determine the optimum expression conditions to produce β_1 protein in *E. coli*. Overcoming challenges in expression of membrane proteins is followed by challenges in maintaining hydrophobic transmembrane segments in aqueous solution, with several viable options available.

As membrane proteins have hydrophobic outward-facing residues, they require the non-polar cell membrane environment (Figure 28) to maintain proper structure and function. Similarly, purified membrane proteins must also be embedded in a membrane-like environment suitable for study with NMR. Since cells are too large to tumble quickly enough for study with NMR as they would have rapid T2 relaxation times,[235] other methods have been developed to maintain a membrane environment while also allowing the protein to tumble at a rate compatible with solution NMR experiments. The use of detergents that form micelles (Figure 28) small enough to rapidly reorient provides a membrane-like environment while still allowing for the protein-micelle complex to rapidly reorient during NMR data collection. Various detergents have been employed to produce

NMR structures of membrane proteins including, but not limited to sodium dodecyl sulfate (SDS), n-dodecylphosphocholine (DPC), and 1,2-diheptanoyl-*sn*-glycero-3-phosphocholine (DHPC)[222] and will be discussed in more detail in section 4.3.3. Newer techniques, like lipid nanodiscs,[236] offer advantages over micelles by preventing aggregation and producing monodispersed protein-lipid complexes, but these systems require expertise in nanodisc preparation. For this study, detergent micelles were utilized due to the wide availability of published methods.[237]

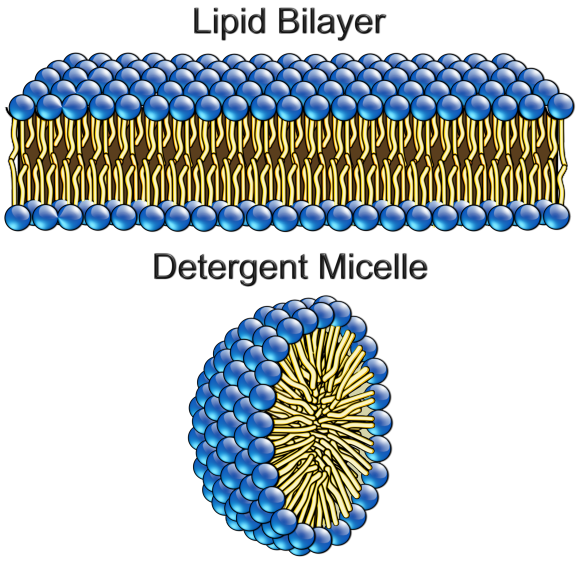


Figure 28. Diagram of Lipid Bilayer and Detergent Micelle

Since β_1 has not been expressed and purified from *E. coli* before, functional assessment of the protein produced is a necessity. Because expression of non-native protein can produce incorrect folding as discussed previously, experiments showing the purified protein can bind known ligands or

exhibit wild type modifications of BK_{Ca} channel function when reconstituted with α subunits are needed to justify the time and expense of NMR experiments. Current techniques for assessing β_1 involve patch-clamp techniques and channel current recordings involving the α-β heterooctomer,[133] however, no functional assays exist for β_1 alone. While several methods were attempted after successfully expressing and purifying the protein, none of the experiments produced results indicating the protein is in a functional, native-like fold.

4.2 Methods

All molecular biology experiments were based on standard protocols from *Molecular Cloning: A Laboratory Manual* by Maniatis *et al*.[238] Examples of experiments performed are given below.

4.2.1 Preparation of β_1 DNA for Subcloning

The *Rattus novergicus* cDNA for the β_1 subunit of the BK_{Ca} channel was obtained from the lab of Dr. Alex Dopico at the University of Tennessee Health Science Center (UTHSC, Memphis, TN). The mRNA sequence information can be accessed through the National Center for Bioinformatics (NCBI) GenBank[239] system using accession code FJ154955.1 (http://www.ncbi.nlm.nih.gov); the DNA was provided in pcDNA3.1 (Invitrogen), a mammalian expression vector. Initial attempts at expression using the *R. novergicus* gene were unsuccessful. To improve expression, the gene was optimized for *E. coli* codon usage using the GeneArt service (Life Technologies) to replace codons rarely used in *E. coli* in the DNA sequence. For the *E. coli* codon optimized DNA sequence as well as the protein sequence of β_1 see Appendix A. The optimized gene was received in a non-expression vector, pMA-T, containing NdeI and XhoI restriction

endonuclease sites before and after the gene, respectively. These restriction enzyme sites made it possible to excise the gene using these enzymes and subclone directly into pET-28a (Novagen, Figure 27) without the need for polymerase chain reaction (PCR) intermediates.

Plasmid propagation was accomplished using the NovaBlue cell line, a specialized *E. coli* cell line known to produce large quantities of plasmids. The pMA-T plasmid DNA containing the β_1 gene was used to transform NovaBlue *E. coli* competent cells (EMD Biosciences) using a standard heat-shock protocol. Transformation was completed by adding 3 µL of 25 ng/µL purified DNA to a sterile vial containing 25 µL of NovaBlue competent cells. The cells and DNA were allowed to incubate on ice for 30 minutes before being placed in a 42˚C water bath for 30 seconds. The vial was then placed back on ice for 5 minutes before adding 250 µL of SOC (super optimal broth with catabolite expression) media (Cellgro) to the vial followed by agitation on an orbital shaker at 37˚C and 250 rpm. After one hour, the tubes were removed from the orbital shaker and 50 µL of the bacterial solution was plated on a sterile 10 mL lysogenic broth (LB) agar plate containing 1 µg/mL ampicillin. The inoculated plate was placed in an incubator at 37˚C overnight. Each colony selected from the plate after incubation was used to inoculate 10 mL of LB media containing 1 µg/mL of ampicillin in a sterile tube. The 10 mL cultures were left overnight in an orbital shaker at 37˚C and 250 rpm. The following morning the cells were pelleted by centrifuging at 8864.5xG for 10 minutes and the plasmid DNA was purified from *E. coli* cultures using a QIAprep Spin Miniprep Kit (Qiagen); manufacturer's instructions were

followed. DNA concentrations were determined using standard ultraviolet spectroscopy protocols.[238] Purified DNA was sequenced by the UTHSC Molecular Resource Center (UTHSC MRC, Memphis, TN) and then translated into the corresponding amino acid sequence using the ExPASy[240] Translate tool (http://web.expasy.org/translate/). The translated sequence was aligned with a translated form of the GenBank sequence using the NCBI Protein BLAST algorithm[241] (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

4.2.2 Excision of β_1 Gene and Ligation Into pET-28a

After sequencing results verified that the plasmid contained the correct sequence, the purified pMA-T DNA was digested using NdeI and XhoI to excise the β_1 gene. The reaction was prepared with 1 µL of NdeI (20,000 U/mL, NEB), 1 µL of XhoI (20,000 U/mL, NEB), 0.5 µL of 10 mg/mL purified bovine serum albumin (BSA, NEB), 2 µL of NEB 10X concentrated React 4 buffer, 5 µL of 289 ng/µL purified pMA-T β_1 DNA, and 10.5 µL of purified water to give a final reaction volume of 20 µL in a sterile microcentrifuge tube. The tube was placed in an incubator at 37˚C for 1.5 hours. Purified pET-28a was digested using the same enzymatic digestion protocol described above using 1µg of plasmid DNA. Both digestion products were purified using a 1% agarose gel made with 40 mM tris, 20 mM acetic acid, and 1 mM EDTA buffer (TAE buffer) at 90 V for 70 minutes using a BioRad PowerPac Basic power supply and an Owl EasyCast B1A mini gel electrophoresis system (Thermo Scientific). Hi-Lo (Bionexus) and MassRuler Low Range (Thermo) DNA ladders were used for estimating DNA fragment sizes. All gels were imaged using the Fotodyne FOTO/Analyst PC

Image Version 5.00. Appropriate DNA bands were excised and purified from TAE gels using the QIAquick Gel Extraction Kit (Qiagen).

Ligation was carried out using a 1:1 molar ratio of digested β_1 DNA and digested pET-28a DNA. T4 DNA ligase (Thermo Scientific), along with the reaction buffer containing ATP, were used for DNA ligation and were allowed to react overnight at room temperature. The following morning 4 µL of the ligation reaction was combined with 25 µL of XL-10 Gold ultra-competent cells (Agilent). The cells were transformed and DNA was propagated and purified as described in section 4.2.1. The presence of the β_1 gene was visually verified by digestion with NdeI and XhoI as described previously in the current section. DNA sequencing by UTHSC MRC was used to verify the sequence.

4.2.3 Polymerase Chain Reaction Amplification of the β_1 Gene

DNA oligonucleotide primers were designed to amplify the β_1 gene for ligation into pE-SUMOpro (Figure 27, Table 8). PCR[242,243] was used to amplify the gene with the appropriate restriction endonuclease recognition sequences. The forward primer added a SfaNI (NEB) restriction site, while the reverse primer added an XhoI restriction site. The use of these two restriction sites added nucleotides to the β_1 gene to produce complementary overhangs to the linearized form of pE-SUMOpro. This allowed for ligation of the gene into the vector. The Q5 High-Fidelity PCR Kit (NEB) was used to perform PCR reactions; manufacturer's instructions were followed for reaction concentrations and PCR thermal cycling programs. The Eppendorf Mastercycler Personal machine was used to carry out all PCR reactions.

Table 8. PCR Primers for Ligation of the β_1 Gene Into pE-SUMOpro
The underlined sequences are the sequences of DNA overhangs from restriction enzyme cleavage sites needed for ligation into pE-SUMOpro linearized by BsaI. The forward primer contains an SfaNI cleavage site, while the reverse primer contains an XhoI cleavage site (shown in grey).

Identifier	DNA Sequence
pE-SUMOpro Forward Primer	CCGCGAACAGGTGGCATCTCTCTAGGTATGGGCAAAA AACTGGTTATGG
pE-SUMOpro Reverse Primer	GTCACCGTACAAGCCTCGAGTTATTTCTGTGCTGCCAG AAC

The vector was prepared for ligation by linearizing with the restriction endonuclease BsaI (NEB). The β_1 PCR product was prepared for ligation by digesting with SfaNI and XhoI and the ligation reaction, bacterial transformation, plasmid purification, and DNA sequence verification were carried out as described in section 4.2.2.

4.2.4 Ligation Independent Cloning of β_1 into pMCSG19 and pMCSG29

The β_1 gene was subcloned into pMCSG19 and pMCSG29[225] (Figure 27) using ligation independent cloning techniques described in detail by Stolls *et al*.[230] Bacterial transformation, plasmid purification, and sequence verification were performed as described in section 4.2.2.

4.2.5 Expression Trials of β_1 in pET-28a, pE-SUMOpro, pMCSG19, and pMCSG29

Purified and sequence verified forms of all four vectors containing the β_1 gene were used to transform BL21 (DE3) and C43 (DE3)[233] *E. coli* competent cells for protein expression. Starter cultures were prepared in a sterile 50 mL

centrifuge tube with 10 mL of sterile LB broth (Fisher), 1 μg/mL of the appropriate antibiotic, and a freshly isolated bacterial colony. The culture was grown overnight for no more than 16 hours in an incubator shaker at 37˚C and 250 rpm. The following morning, 50 mL of sterile LB broth were prepared with 1 μg/mL of antibiotic in a sterile 250 mL flask and warmed to 37˚C before adding 1 mL of the 10 mL starter culture. The inoculated 250 mL flask was placed back in the incubator shaker at 37˚C and 250 rpm and the optical density (OD) was measured at 600 nm using an Agilent 8453 UV-Vis Spectrophotometer every hour until reaching an OD of 1.0. Protein expression was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) from a 1 M stock to a final concentration of 0.5 mM. Expression was carried out at 18˚C in some cases; cultures were grown as described, but switched to an 18˚C refrigerated shaker after induction of expression. A 1 mL sample of the bacterial culture was taken before inducing expression with IPTG and the sample was centrifuged at 16,000xG to pellet cells. The supernatant was decanted and the cell pellet was stored at -20˚C. Additional samples were taken every hour or two hours and the number of cells in each sample was normalized using OD measurements. All samples were thawed at room temperature and re-suspended in 300 μL of lysis buffer (75 mM trizma base, 0.3 M NaCl, 0.2 mM EDTA to reduce proteolytic cleavage, and 10 μM butylated hyroxytoluene to act as a free-radical scavenger) with protease inhibitors. Roche complete protease inhibitor tablets were used and were prepared by dissolving one tablet in 50 mL of lysis buffer.

Cell lysis was accomplished using the Biologics Ultrasonic Homogenizer 150 V/T with a titanium microtip. Sonicator power was set to 30% of maximum and a 50% pulse program (5 seconds on, 5 seconds off) was used for 3 minutes per sample with samples kept on ice. Whole cell protein samples were prepared by removing 25 µL of each sample into a sterile microcentrifuge tube and 25 µL of Laemmli sample buffer (BioRad, 65.8 mM Tris-HCl, pH 6.8, 2.1% SDS, 26.3% (w/v) glycerol, 0.01% bromophenol blue, 355 mM β-mercaptoethanol) was added to each tube. The remaining sonicated sample was then centrifuged at 16,000xG under refrigeration to 4˚C for 10 minutes to pellet insoluble cell debris. Cell lysate samples were prepared by pipetting 25 µL of the supernatant from the centrifuged samples and combining it with 25 µL of Laemmli sample buffer. Both sets of samples were placed on a rocker at room temperature for 30 minutes.

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)[244] was used to analyze expression time course samples. Fifteen µL of each time point's whole cell and cell lysate samples were loaded into Mini-Protean TGX AnykD 15 well precast gels (BioRad) along with Precision Plus Protein Kaleidoscope (BioRad) protein ladder. Tris/glycine/SDS running buffer (BioRad; 25 mM Tris, 192 mM glycine, 0.1% SDS, pH 8.3) was added to the Mini-Protean Tetra Cell (BioRad) electrophoresis cell and all gels were run at 200 V for 33 minutes. The gels were removed from cassettes and placed in a gel fixing solution containing 50% methanol and 10% acetic acid for 1 hour. The gels were then rinsed in water for 10 minutes and stained in Bio-Safe Coomasie

(BioRad) for one hour and de-stained in water for a minimum of one hour. All gels were imaged using the Fotodyne FOTO/Analyst PC Image Version 5.00.

4.2.6 Large-Scale Expression of β_1 in *E. coli*

Large-scale expression of β_1 protein was carried out in 4 L sterile flasks using 1 L of sterile LB broth. Fifty mL starter cultures were prepared as described in section 4.2.5 using pET-28a and pE-SUMOpro β_1 constructs. Starter cultures were used to inoculate 1 L of sterile LB, which was pre-warmed to 37˚C in an incubator shaker. Cultures were grown to an OD of 1.0 before inducing protein expression by adding IPTG to a final concentration of 0.5 mM followed by 4 hours incubation at 37˚C and 250 rpm. After 4 hours, the flasks were removed from the incubator shaker and placed on ice. Cells in the 1 L cultures were divided into 250 mL portions and pelleted by centrifuging in a Sorvall RC-5B refrigerated superspeed centrifuge with a GSA rotor at 5855xG for 30 minutes. Cell pellets that were not immediately used for purification were stored at -20˚C.

4.2.7 Purification of β_1 from Inclusion Bodies

Frozen cell pellets from large-scale expression were removed from -20˚C storage and allowed to thaw at room temperature for 10 minutes. Pellets were re-suspended in 20 mL of lysis buffer containing protease inhibitors and lysed as described in section 4.2.5. Centrifuging at 8800xG for 20 minutes pelleted cell debris, including protein inclusion bodies. The supernatant was discarded and the pellet was solubilized with 200 mL of lysis buffer + 0.5% SDS in a sterile 250 mL flask in an incubator shaker at 37˚C. After one hour, the solution was centrifuged at 8800xG for 10 minutes to pellet any insoluble mass and the supernatant was transferred to a sterile flask. The supernatant was filtered using

a Millex GP 0.22 µm syringe-driven filter (Millipore) and a sterile syringe. Before solubilized protein was applied to the column, a loading sample was prepared by collecting 25 µL of solubilized protein and combining it with 25 µL of Laemmli buffer, as described in section 4.2.5. The filtered supernatant was applied to a HisTrap FF 5 mL Ni-NTA (GE Healthcare) IMAC column at a 5 mL/min flow rate using a Gilson MiniPuls3 peristaltic pump. An additional sample of the protein solution was taken after exposure to the column as described above and is referred to as the flow through sample. The loaded column was connected to the BioRad BioLogic LP chromatography system and all UV data was recorded using LP Data View MS Software Version 2.00 (BioRad). All purifications used a flow rate of 5 mL/min and UV data was recorded at 280 nm.

To purify protein, the column was first treated with wash buffer (40 mM HEPES, 300 mM NaCl, 10 µM butylated hydroxytoluene, 0.5% SDS, pH 7.2) for 10 minutes. On-column detergent exchange was carried out using rinse buffer (50 mM Na_2HPO_4, 300 mM NaCl, and 0.25% Fos-Choline-12, Sol-Grade (DPC, Affymetrix)) for 10 minutes. The purified protein was eluted from the column by washing the column with elution buffer (250 mM imidazole, 300 mM NaCl, 0.25% DPC or other detergents, pH 4) for 10 minutes. The pH of elution fractions was raised immediately after collection to 7.4 using 1 M Trizma base. Elution buffer was made using 0.25% SDS, 0.2% N-lauroylsarcosine (sarcosyl), or 0.2% 1,2-diheptanoyl-*sn*-glycero-3-phosphocholine (DHPC) in place of 0.2% DPC to elute the protein in different detergent micelles for later study. All fractions were

collected in 5 mL increments, one per minute, using the BioRad Model 2110 fraction collector.

Samples of each fraction were prepared by transferring 25 µL of each fraction to a sterile microcentrifuge tube and combining it with 25 µL of Laemmli buffer for analysis via SDS-PAGE, as described in section 4.2.5. Fractions containing purified protein were pooled and concentrated using Amicon Ultra-15 centrifugal filter units with Ultracel-3 membrane (Millipore) by centrifuging at 8800xG in 20 minute increments until the desired volume was obtained. The elution buffer was exchanged to phosphate buffered saline (PBS; 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl, pH 7.4). Protein concentration was determined using the BCA Assay Kit (Thermo) following manufacturer's instructions. Excel (Microsoft) was used for standard curve fitting of BCA assay data and concentration calculations.

4.2.8 Binding Competition Assay of β_1 Using Fluorescent Estradiol

Fluorescently labeled BSA covalently linked to β-estradiol and fluorescein was used to measure binding of estradiol to β_1 protein purified from *E. coli*. A 100 µM stock of β-estradiol 6-(O-carboxy-methyl)oxime: BSA fluorescein iso-thiocyanate conjugate (BSA-E-F; Sigma, approximate molecular weight for lot purchased was 70,900)[245] was prepared by dissolving 0.354 mg in 50 µL of purified water. A 1 mM stock of non-labeled estradiol was prepared by dissolving 0.272 mg of estradiol in 1 mL of ethanol. Several serial dilutions were made (Table 9) to test 0.5, 1, 2, and 5 µM concentrations of BSA-E-F in the presence or absence of β-estradiol and β_1 protein purified from BL21 (DE3) cells in 0.2%

SDS, DHPC, and sarcosyl detergent micelles (section 4.2.7). A 96-well His-select nickel coated plate (Sigma) was used to bind His tagged protein to the surface of the plate. Ethanol was added to solutions that did not contain β-estradiol, which was first dissolved in ethanol, to keep all solutions of equal composition. All solutions to which β-estradiol was added had a final concentration of 5 μM estradiol.

Table 9. Solutions Prepared for Fluorescent Binding Competition Assay

Final Concentration of BSA-E-F (μM)	Volume of 100μM BSA-E-F Added (μL)	Volume of PBS (μL)	Volume of EtOH (μL)	Volume of 1mM Estradiol Added (μL)
0	0	99.5	0.5	0
0.5	0.5	99	0.5	0
1	1	98.5	0.5	0
2	2	97.5	0.5	0
5	5	94.5	0.5	0
5	5	94.5	0	0.5
0	0	99.5	0	0.5

Protein binding was accomplished by adding 200 μL of PBS containing 5 μg/mL protein purified into detergent micelles to each well. The protein was allowed to incubate in the wells for 1 hour and the remaining solution was decanted. The wells were washed three times each with 300 μL of PBS containing the appropriate detergent. One hundred μL of each solution in Table 9 was pipetted into each well in triplicate and allowed to incubate for one hour. Due to a low yield of protein purified in sarcosyl micelles, only the 5 μM

concentrations with and without β-estradiol were used. After incubation, the wells were washed three times with 300 µL of PBS containing the appropriate detergent. Fluorescence was measured at 521 nm with a 6 minute total run time taking fluorescence measurements every 2 minutes on a Synergy 2 plate reader (BioTek) and data was acquired on a computer running the Gen5 software package (BioTek). Blanks of 5 µM BSA-E-F with no protein and PBS only were used as controls, also in triplicate. The average and standard deviation of each detergent and BSA-E-F concentration combination were calculated using Excel (Microsoft).

4.2.9 Surface Plasmon Resonance Binding Studies with Purified β_1

A nickel-coated surface plasmon resonance (SPR) chip was used to measure binding of β_1 and ligands. The NiP SPR sensordisc (Xantec Bioanalytics) was cleaned with purified water and dried completely using nitrogen gas. The disk was loaded into the cell of the Autolab ESPRIT SPR instrument (Eco Chemie) and data was acquired on a computer running the accompanying Data Acquisition software version 4.2.2. The chip and cells were first rinsed with PBS in both channels before applying protein solution. Fifty µL of 5 µM solutions of purified protein were applied to both channels of the SPR cell and allowed to associate with the nickel-coated surface for 10 minutes with pump-controlled mixing. The cell was then washed again with PBS before adding 50 µL of a 5 µM solution of β-estradiol 6-(O-carboxy-methyl)oxime: BSA (BSA-E, Sigma) to one channel and 50 µL of a 5 µM solution of BSA (Thermo) to the other channel. The solutions were left in the cell for 10 minutes with mixing before being washed with PBS. Finally, 50 µL of 50 µM β-estradiol was loaded into each channel and

allowed to incubate with mixing for 10 minutes followed by a wash with PBS. Binding measurements were acquired throughout all solution applications and washes.

4.2.10 Site Directed Mutagenesis of β_1 for Cholane Binding Studies

Custom DNA oligo primers were ordered from Operon to create a β_1 T169A mutant in the pET-28a construct. The sequences of the primers are shown in Table 10. Primers were reconstituted in purified water to a concentration of 100 μM. Mutagenesis was performed using a QuickChange Lightning Site-Directed Mutagenesis Kit (Agilent) and manufacturer instructions were followed. An Eppendorf Mastercycler Personal PCR machine was used to perform the reaction in duplicate following the manufacturer's PCR thermal cycling sequence. The PCR products were transformed into XL1 Blue competent cells as described in section 4.2.2. Plasmid propagation, purification, and sequence verification were also performed as described in section 4.2.2. Expression trials and large-scale growth were performed as described in sections 4.2.5 and 4.2.6, respectively.

Table 10. Primers to Create β_1 T169A Mutant

Identifier	DNA Sequence
T169A Forward Primer	CCTTTCTGCTGGCCGGTGGTCTGCT
T169A Reverse Primer	AGCAGACCACCGGCCAGCAGAAAGG

4.2.11 Binding Studies with Lithocholic Acid-Linked Magnetic Beads

The synthesis of magnetic beads covalently linked to lithocholic acid (LCA) was completed by Lisa Chaney, an undergraduate student participating in the Research Experience for Undergraduates program in the summer of 2013. The unpublished synthesis is described for context.

BcMag Amine-Terminated beads (Bioclone) were rinsed with coupling buffer (10 mM K_2HPO_4, 0.15 M NaCl, pH 5.5) three times as instructed by the manufacturer. Supernatant was removed by placing microcentrifuge tubes containing the magnetic beads on the supplied magnetic separator and excess solution was decanted. The beads were resuspended in coupling buffer and combined with LCA solution (0.0346 g lithocholic acid (Sigma), 1.75 mL anhydrous 1,4-dioxane (Sigma), 0.75 mL purified water) and a coupling agent solution (0.0346g 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide, 25 mL purified water) and left at room temperature on a rocker. After 24 hours, the beads were rinsed with purified water and storage buffer (0.303 g tris base, 7.305 g NaCl, 0.09 g EDTA, 0.25 g sodium azide, pH 6.99) three times. The beads were resuspended in storage buffer and stored at 4˚C until needed.

For functional assessment, the magnetic separator supplied by the manufacturer was used to separate the beads and the storage buffer was decanted. The beads were washed three times with 1 mL of PBS and then resuspended in 2 mL of PBS. One mL aliquots of the PBS bead suspension were prepared in separate microcentrifuge tubes and 250 µL of 16 µg/mL WT β_1 protein was added to one and 250 µL of 16 µg/mL T169A β_1 was added to the other. A control tube was prepared by combining 250 µL of 16 µg/mL WT β_1

protein with unreacted magnetic beads. The tubes were placed on a rocker. After 3 hours, the tubes were removed and the magnetic separator was used to separate the beads from the solution, which was decanted. The beads were then washed three times with PBS and 100 µL of 2x Laemmli buffer (BioRad) was added to each tube along with 100 µL of PBS to achieve 1x concentration of Laemmli. The tubes were inverted to mix and then placed on the rocker for 3 minutes. The first elution sample (E1) was prepared by transferring 60 µL of the buffer to a new tube and the beads were placed back on the rocker overnight. The following morning a second elution sample (E2) was prepared in the same way and samples were analyzed using SDS-PAGE as described in section 4.2.5.

4.3 Results and Discussion

4.3.1 PCR and Vector Ligation

After excision from the pMA-T vector received from GeneArt (Figure 29 A), the β_1 gene was ligated into pET-28a. An enzymatic digestion of the pET-28a β_1 construct run on a 1% agarose TAE gel with the MassRuler DNA ladder showed the expected β_1 DNA fragment at ~600 bp (Figure 29 B). Since the pET-28a construct including the His tag and β_1 gene produces a 212 amino acid protein, the expected fragment is 636 bp and is in agreement with the agarose gel. Three DNA samples were purified from three separate colonies transformed using the ligation reaction of the β_1 gene and pET-28a. All three DNA samples were digested, and showed a DNA fragment between the 500 bp and 750 bp markers in the Hi-Lo DNA ladder (Bionexus), visually confirming the presence of the β_1 gene. DNA sequencing results confirmed the presence of the β_1 gene in the pET-28a vector.

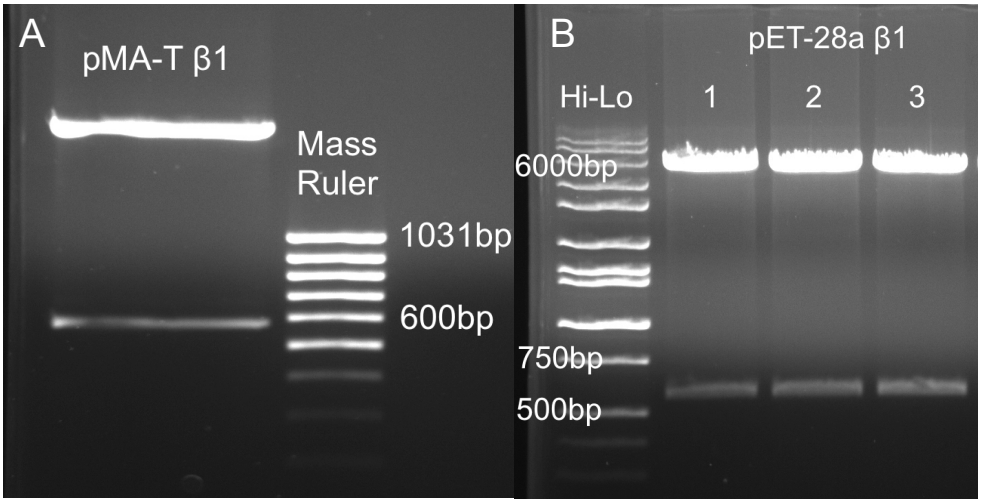


Figure 29. pMA-T and pET-28a β_1 Construct Digestion with NdeI and XhoI
A) Double digestion of pMA-T vector containing the β_1 gene. B) Double digestion of pET-28a vector containing the β_1 gene. Three bacterial colonies were selected after ligation of the β_1 gene into pET-28a. All three colonies contained the correct gene insert.

Agarose gel electrophoresis and DNA sequencing also confirmed PCR amplification of the β_1 gene for LIC insertion into pMCSG19 and pMCSG29. The LIC method proved less reliable than traditional subcloning at first, so 20 colonies were selected after insertion of the β_1 gene into pMCSG19 and purified by miniprep before digesting with NdeI and HindIII (New England Biolabs, Figure 30 A). Due to the presence of the MBP fusion partner in both pMCSG19 and pMCSG29, the fragments are much larger than the fragments from pET-28a, which contains only a His tag and the β_1 gene. DNA fragments containing the β_1 and MBP fusion partner gene from pMCSG19 were expected at ~2000 bp, while DNA fragments without the β_1 gene were expected at ~1400 bp. The NEB Broad Range DNA ladder was used for pMCSG19 digests while pMCSG29 digests

used Hi-Lo (Bionexus) due to availability issues. pMCSG29 was also digested (Figure 30 B), but XbaI and BamHI restriction enzymes (New England Biolabs) were used instead due to differences in available DNA restriction enzyme sites between vectors.

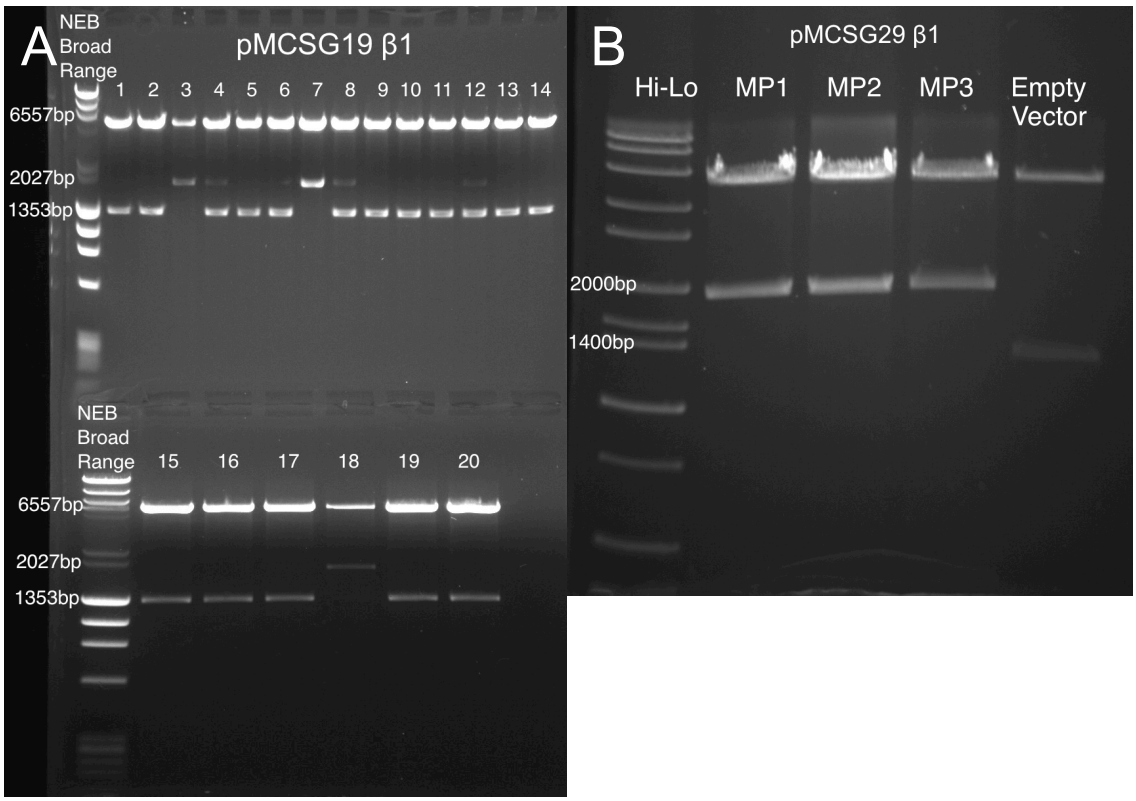


Figure 30. pMSCG19 and pMCSG29 β_1 Construct Double Digest
A) Double enzymatic digestion of the pMCSG19 β_1 construct by NdeI and HindIII. Samples containing the β_1 gene insert show bands at ~2000 bp, while samples without the insert show bands at ~1400 bp. Samples 3, 7, and 18 clearly showed presence of the β_1 insert, while samples 4, 8, and 12 faintly show the insert and possible contamination from empty vector. B) All three miniprep samples digested with XbaI and BamHI showed the presence of the β_1 insert while the digested empty vector showed the expected band at ~1400 bp.

In Figure 30 A, samples 3, 7, and 18 clearly show the presence of the β_1 insert, while samples 4, 8, and 12 show a faint band indicating the presence of the insert, but also show a band without insert. This could have been caused by collecting more than one bacterial colony from the agar plate during colony picking or could have been a result of uptake of more than one plasmid by the bacteria in the selected colony. These contaminated samples were not verified by sequencing, but the pure samples were used to confirm the presence of the β_1 insert. Figure 30 B shows the LIC method worked more reliably for pMCSG29 with only three colonies being selected for digestion and all three miniprep samples containing the correct sequence, which was verified with DNA sequencing as described in section 4.2.2.

The difference in success between LIC methods in pMCSG19 and pMCSG29 could be due to the difference in the nucleotides used in the LIC procedure described by Stols *et al*.[230,225] pMCSG19 uses adenine and thymine base pairing to produce LIC overhangs in the DNA fragments, while pMCSG29 uses guanine and cytosine base pairing. Commercially available kits from Novagen utilize the GC base pairing method used in pMCSG29, but no commercially available kits could be found utilizing the AT base pairing in pMCSG19. Problems also could have arisen from the use of 10 mM nucleotide stocks used with pMCSG19 vector and insert preparation, while 100 mM nucleotide stocks were used for pMCSG29. Glycerol and other stabilizing agents in the less concentrated stocks, which required larger volumes to complete the reactions, could have negatively impacted the function of T4 DNA polymerase,

interfering with the reaction necessary to produce complementary LIC overhangs on the vector and insert.

4.3.2 Expressions of pET-28a, pE-SUMOpro, pMCSG19, and pMCSG29 β_1 Constructs

In this study, four bacterial expression vectors were used to determine the optimum expression plasmid for the β_1 transmembrane protein in *E. coli* (Figure 27). The first vector examined was pET-28a (Novagen), a basic *E. coli* expression plasmid containing a 6x-His tag for IMAC purification, a thrombin protease cleavage site, and the kanamycin antibiotic resistance gene for a selection marker. The second vector used was pE-SUMOpro Amp (LifeSensors), which contains a small ubiquitin-like modifier (SUMO) protein fusion partner for solubility enhancement[232], a 6x-histidine tag for IMAC purification, and a SUMO Protease 2 cleavage site to remove the SUMO fusion protein after expression and purification. The SUMO vector also contains a gene for ampicillin antibiotic resistance as a selection marker. Both pET-28a and pE-SUMO use a multiple cloning site for gene insertion.

The fusion protein in the pET-28a construct has a molecular weight of 25 kiloDaltons (kDa), but appears in SDS-PAGE gels around 20 kDa caused by the hydrophobic nature of transmembrane segments binding more SDS than cytosolic proteins.[246] pET-28a produced visible overexpression in the whole cell samples from BL21 cells at 18 and 37˚C (Figure 31 A, C, and D), but failed to show any protein in C43 cells at 37˚C (Figure 31 B), so 18˚C growth was not pursued. This construct is the favored construct for solution NMR, attributable to the lack of a large fusion partner that requires post-purification cleavage and

reduces protein yields. Samples labeled “L” in Figure 31 represent the lysate samples, while “W” samples represent whole cell samples whose collection and preparation were described in section 4.2.5. Since the protein being expressed is a membrane protein, it is not expected to express in lysate samples, which only contain soluble cytosolic proteins. Whole cell samples contain the cytosolic proteins, cell membranes, cell debris, and any protein aggregates or inclusion bodies. The stability of the protein expressed at 18˚C, which can clearly be seen through 48 hours of sampling, could indicate inclusion body formation. Trials using the pET-28a vector revealed that expression in the BL21 cell line was favored over C43 cells and that expression at 18˚C, which can improve expression yields of membrane proteins[182], was unsuccessful in improving protein yield.

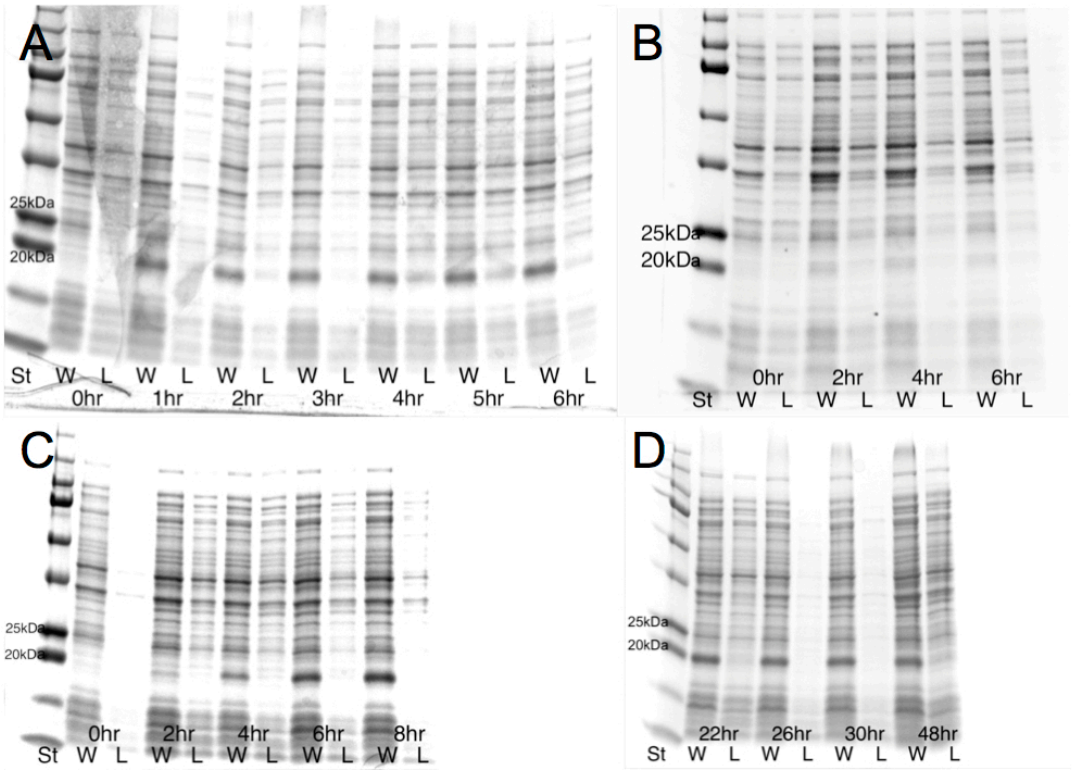


Figure 31. Expression Trials of the pET-28a β_1 Construct
Samples labeled “L” represent the lysate samples, while “W” samples represent whole cell samples. Times displayed represent time passed since induction with IPTG. The Kaleidoscope protein standard was used for all gels and is denoted as “St” A) Expression in BL21 cells at 37˚C indicated overexpression at the 1 hour time point, which remained stable, and is evidenced by the band appearing around 20kDa. B) Expression in C43 cells showed minimal difference in the band at 20 kDa between pre and post-induction samples. C) Expression in BL21 cells at 18˚C first 8 hours of time points. Expression is visible after 2 hours, but is visibly overexpressed after 6 hours. D) Time points 22-48 hours after induction for BL21 cells at 18˚C. Protein remains stable over the extended sampling period.

The SUMO-β_1 fusion protein was expected at ~35 kDa if overexpression were observed owing to the presence of the 12-15 kDa SUMO fusion partner. Figure 32 shows the trial expression of the SUMO-β_1 fusion in BL21 cells at 37˚C. There was no discernable difference in expression levels for pre-induction 0 hour time points compared to post-induction 2, 4, and 6 hour time points indicating there was no overexpression of protein for this construct, as evidenced by the lack of a band appearing in 2, 4, and 6 hour samples that is not also seen in the 0 hour samples. The pE-SUMOpro construct produced no overexpression of β_1-SUMO fusion protein in BL21 or C43 cells in expression trials at 37˚C. Because of the lack of expression in both cell lines, the pE-SUMOpro construct was not pursued any further.

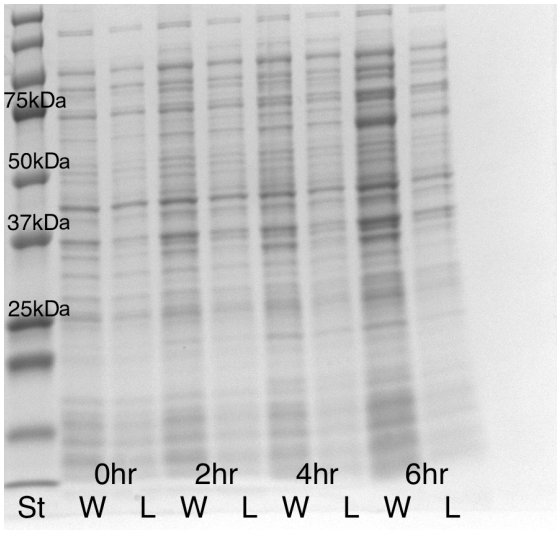


Figure 32. Expression Trial of pE-SUMOpro β_1 Construct
Expression trials of SUMO-β_1 fusion protein failed to show overexpression. The overexpressed protein was expected around 36 kDa.

Expression from the pMCSG19 vector was much higher, as indicated by the large, dark band in Figure 33A in the 2 hour time point “W” sample between the 50 kDa and 75 kDa markers that does not appear in the 0 hour time point. The size of this band could be attributed to the large mass of MBP, which has a mass of 42 kDa, and not overall greater quantity of protein compared to the pET-28a β_1 expression profile. The presence of the band only in the whole cell sample could indicate protein aggregation, but could also be membrane embedded protein. The MBP fusion vectors, pMCSG19 and pMCSG29, both demonstrated overexpression only in C43 cells. pMCSG29 did not express as well as pMCSG19 and was not pursued any further due to the added complication of fusion partner cleavage after purification, but no added benefit of overexpression compared to pET-28a.

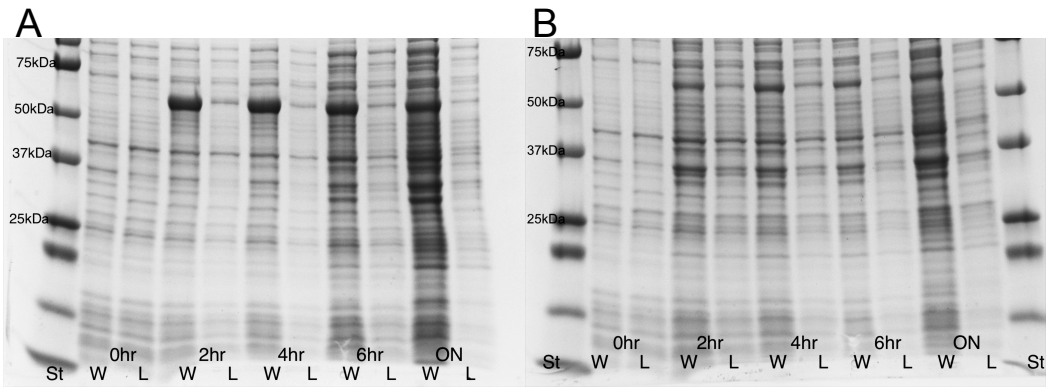


Figure 33. Expression Trials of pMCSG19 and pMCSG29 β_1 Constructs
A) pMCSG19 β_1 expression in C43 cells indicated good overexpression of protein with a band around 60 kDa, but possibly as insoluble aggregates as was seen with pET-28a. B) pMCSG29 β_1 expression did show overexpression, but not as much as pMCSG19.

After comparison of all expression trials, the pET-28a β_1 construct was selected for large-scale expression and purification using IMAC. pET-28a was chosen over pMCSG19 for purification because they both showed expression in the whole cell fraction, but pET-28a does not require post-purification enzymatic digestion to remove a fusion partner. In pMCSG19, the large MBP fusion tag would require digestion to obtain a protein of suitable size for solution NMR study and the digestion step typically results in reduced yield due to additional purification steps required after digestion.

4.3.3 Protein Purification and Solubilization in Detergent Micelles

Four detergents were used for sample preparation: SDS, DPC, DHPC, and N-lauroylsarcosine (sarcosyl). All of these detergents produce micelles at or above concentrations of 0.2% m/v. These detergents offer various lipid head groups and alkyl chain lengths and were selected to provide a range of micelle environments to determine optimum conditions for β_1 stabilization (Figure 34). SDS is an anionic detergent that can denature proteins in high concentration, but has allowed for quality NMR spectra collection for some proteins.[247,248] Sarcosyl is also an anionic detergent, but is less commonly used. It has been used historically to solubilize membrane proteins,[249] but has also been used to successfully stabilize membrane proteins.[250] DPC and DHPC are both zwitterionic detergents and have both been used to characterize many membrane protein structures by solution NMR.[248]

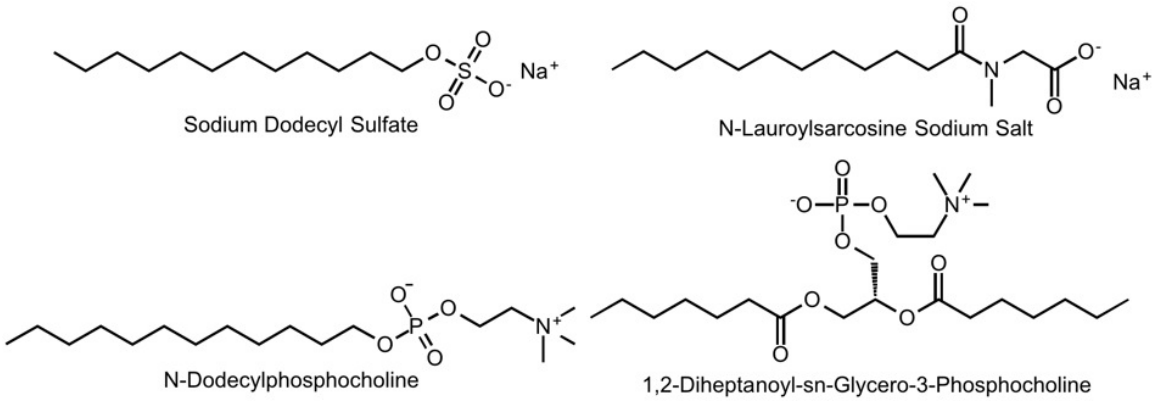


Figure 34. Detergents Used to Stabilize Transmembrane Proteins

Traditional IMAC purification without SDS in the lysis buffer requires a column wash containing 10-50 mM imidazole to remove weakly bound proteins before eluting the purified protein. However, with SDS in the lysis buffer, weak binding proteins did not readily adhere to the column. Recombinant β_1 protein adhered beyond normal IMAC elution requirements of 250-500 mM imidazole after the on-column detergent exchange steps described in section 4.2.7. The elution buffer required a pH of 4 to remove protein from the column by protonating all histidine residues in the His tag, which have a pK_a around 6.[251] The pH was immediately raised to 7.4 following elution to minimize structural issues caused by the low pH environment.[252] Solubilization of the lysed cell pellet described in section 4.2.7 proved to be a very effective method to extract a large amount of protein with high purity (Figure 35), negating the need for a secondary purification step such as anion exchange chromatography.

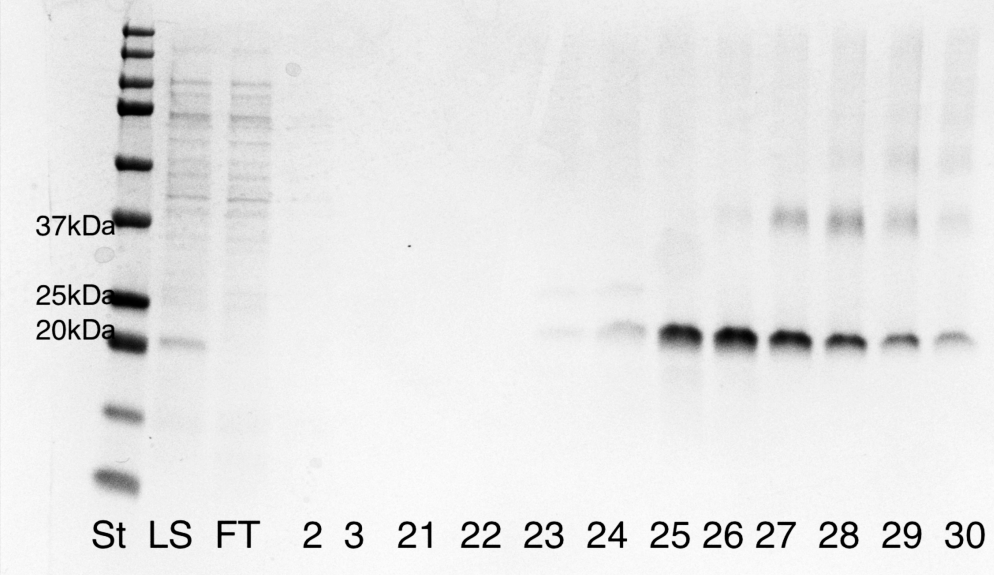


Figure 35. SDS-PAGE Analysis of β_1 Purification into DPC Micelles
The lane labeled “St” is the protein standard for approximating molecular weights. “LS” is the load sample and “FT” is the flow through sample described in section 4.2.7. The numbers represent the fraction collected with the total number of fractions being 30. Fractions 1-10 are column wash fractions, 11-20 are detergent exchange fractions, and 21-30 are elution fractions.

The loading sample and flow through samples in Figure 35 show the solubilized protein, which appears around 20 kDa, was completely absorbed onto the IMAC column since it does not appear in the flow through sample taken after the solution was applied to the column. Column wash fractions 2 and 3 show that very little protein is weakly bound to the column. This is partly due to SDS interfering with binding of weakly attracted proteins and also due to the purification coming from the insoluble cell pellet and not the cell lysate, which contains most of the known contaminating proteins with IMAC affinity in *E. coli*.[253] The faint band that appears around 40 kDa is likely a dimerized aggregate of the solubilized protein. Samples without the band were concentrated separately and after concentration the band appeared in these samples as well. This could prove

to be a significant problem in any future NMR experiments since higher concentrations of protein are required and appearance of the dimer appears to be concentration dependent. This could possibly be avoided by using higher concentrations of detergent or carefully controlling the concentration to promote monodispersion of protein in micelles.[254] Purification using IMAC led to successful purification and on-column detergent exchange into detergents other than SDS, which was used to solubilize protein initially.

4.3.4 Functional Assessment of Purified β_1 Protein

Once purified, it must be determined if protein exhibits native-like behavior, and therefore proper folding, through functional studies. Since β_1 had not been expressed in and purified from *E. coli* before, development of a functional assay proved to be a significant hurdle. With β_1-specific ligands limited to hydrophobic compounds that readily partition into hydrophobic membrane protein-stabilizing environments,[255] measuring the binding of ligands alone would not suffice, as non-specific binding of ligands would be high in the presence of detergent micelles. To deal with this problem several approaches were conceived, all of which center on β_1 ligands covalently linked to larger macromolecules in an effort to reduce non-specific binding.

In the fluorescently labeled estradiol assay described in section 4.2.8, the average signals did appear to increase with increasing concentration of BSA-E-F; however, competition with unlabeled estradiol did not substantially diminish the signal. Additionally, measurements from wells containing no fluorescein were 3.5-4 standardized fluorescence units (SFU), about half of the total signal detected at the highest concentration of BSA-E-F indicating the signal-to-noise

ratio is very low(Table 11). This was likely due to the limited path length of the 96-well plate since fluorescence measured by the plate reader was only coming from ligand bound to protein immobilized at the bottom of the nickel coated plate. With a very small path length, differences in fluorescence between samples were difficult to discern from background. Additionally, the estradiol could be partitioning into the membrane due to the hydrophobic nature of estradiol and the hydrophobic environment provided by the micelle preventing any competition with BSA-E-F. The fluorescently linked β-estradiol assay described in section 4.2.8 failed to produce significant differences between samples with and without estradiol, and thus could not confirm functional properly-folded protein had been purified.

Table 11. β_1 Fluorescein Assay Results
Average and standard deviation calculations from BSA-E-F fluorescence assay of purified β_1 protein incubated with fluorescently labeled BA conjugated to estradiol. Averages of fluorescence readings failed to show a significant difference between different concentrations of BSA-E-F.

	SDS		DHPC		Sarcosyl	
	Average	St Dev	Average	St Dev	Average	St Dev
0 BSA-E	3.56	0.53	3.89	0.33		
0.5 μM BSA-E-F	7.00	0.71	7.67	1.22		
1 μM BSA-E-F	7.44	0.53	7.22	0.44		
2 μM BSA-E-F	8.00	0.50	7.44	0.53		
5 μM BSA-E-F	8.89	0.78	8.56	0.53	17.25	6.18
5 μM BSA-E-F + 5 μM Estradiol	8.67	1.00	12.5	5.59	8.33	1.15
5 μM Estradiol	4.00	0.50	4.00	0.00		

Since SPR can be used to detect protein interactions on the surface of the SPR chip,[256] it was selected to study the interaction of β_1 with ligands. However, since none of the β_1 ligands have a mass large enough to detect with SPR, BSA covalently linked to estradiol was used to measure binding with BSA being used as a control. In theory, if the β_1 protein were correctly folded, the BSA-E would bind and produce a measurable change on the surface of the SPR chip while the normal BSA would not. Addition of estradiol should displace the bound BSA-E while having no effect in the cell containing β_1 and BSA.

Figure 36 shows the association of His tagged β_1 with the Ni-NTA SPR chip. Even though equal amounts of protein solution were put in each channel, a difference of approximately 300 m˚ was detected between the two channels on

the SPR chip. This could be due to reuse of the chip causing degradation of the Ni-NTA surface. However, in Figure 37 after BSA and BSA-E are added, a mass change of ~300 m˚ was measured in channel 1 where BSA-E was added compared to BSA alone even though channel 1 contained a smaller amount of bound protein. Finally, in Figure 38 estradiol was added in 10-fold excess, but failed to show a significant difference in the mass lost between channel 1 and channel 2. Since channel 1 showed a larger mass gain when BSA-E was added, it was expected that the addition of estradiol would compete with BSA-E and the mass gain from Figure 37 would be measured as a mass loss in Figure 38. While the surface plasmon resonance experiments described in section 4.2.9 produced more promising results than the fluorescence assay, they still failed to show reversible binding of the BSA-E substrate when a 10-fold excess of β-estradiol was applied.

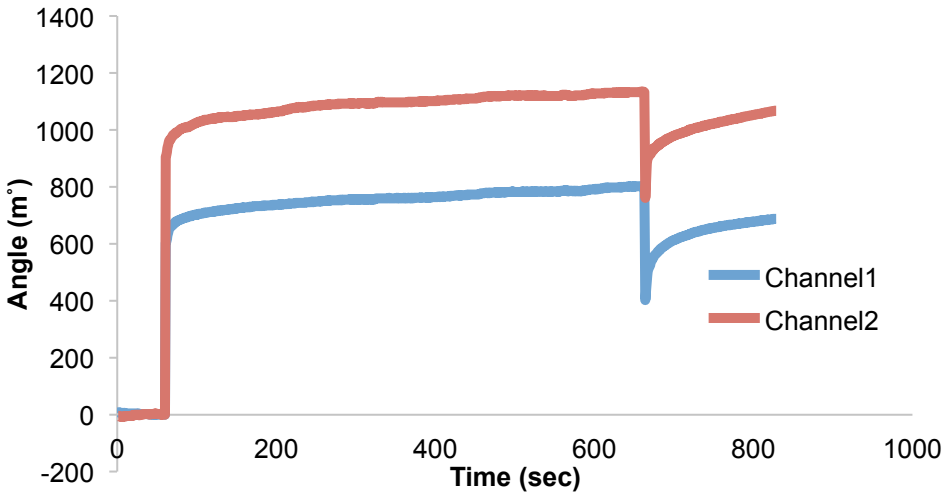


Figure 36. Binding of β_1 in DPC Micelles to SPR Chip

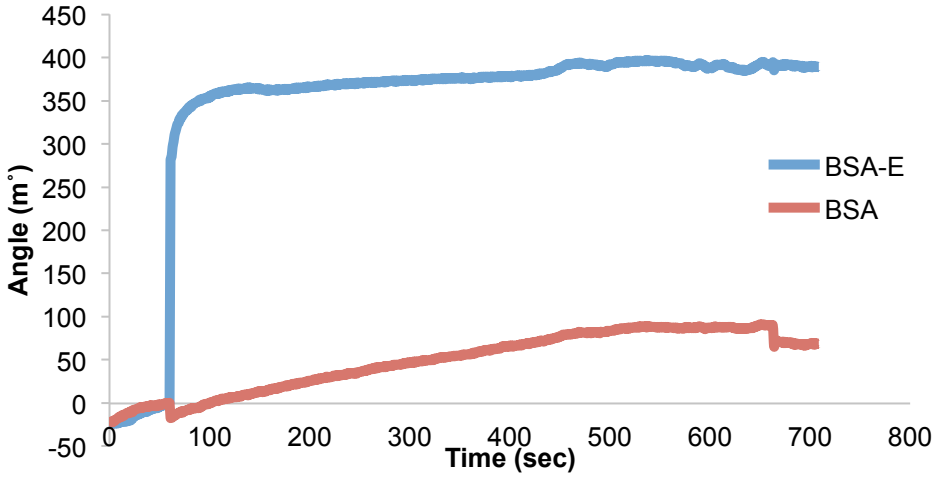


Figure 37. Ligand Binding to β_1 in DPC Micelles on SPR Chip

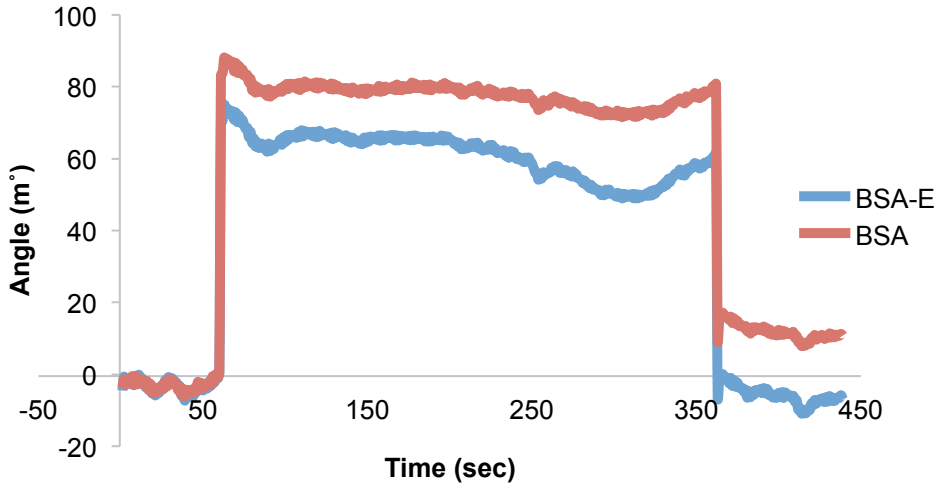


Figure 38. Ligand Competition for β_1 in DPC Micelles on SPR Chip

The failure to see a measurable mass change after addition of estradiol could be due to the BSA-E not binding to β_1 on the surface of the chip and instead binding to micelles or interacting with the SPR chip surface.[257] The data could also imply that the purified protein does not posses native structure since the ligand was unable to be removed through competition by estradiol. However, there could also be an issue with using estradiol for competition since the hydrophobic nature of the molecule will encourage it to partitioning into the detergent micelles present on the chip surface, which may be preferred over occupied ligand binding sites on β_1. Additionally, since estradiol may require the presence of the α subunit for binding[143], it may not be possible for binding to occur with this ligand. As a result of these issues, SPR and estradiol were abandoned as functional assessment tools.

The final method used to assess functionality of β_1 protein used a similar concept as the first two by covalently linking a known hydrophobic ligand to a large non-hydrophobic molecule. However, this method differed in that it used LCA linked to magnetic beads instead of estradiol linked to BSA. The cholane steroid-recognition site containing T169 is well characterized and has a known mutant (T169A) that does not exhibit the wild type LCA response in patch-clamp experiments using the T169A mutant.[65] Because of a better understanding of LCA-β_1 interaction and the use of a mutant as a negative control, this method was pursued to overcome the issues with using non-linked hydrophobic ligands suspected of partitioning into micelles.

In experiments with the magnetic beads covalently linked to LCA, both WT and T169A protein was bound to the beads by comparing the samples before exposure to the LCA beads (Sa) to samples after exposure to the LCA beads (Su) in Figure 39. Additionally, both WT and T169A proteins eluted in E1 and E2 elution samples around 20kDa. While it is difficult to compare Sa and Su to E1 and E2 due to differences in sample volumes and therefore concentrations of proteins in the samples, the Sa and Su samples from T169A in Figure 39 around the 20kDa marker show a greater uptake of β_1 compared to other proteins in the sample. However, since there was no difference in protein bound to the beads in WT compared to T169A, this method also failed to demonstrate the functionality of β_1 protein expressed in *E. coli*. The non-specific binding of both proteins could be due to LCA still interacting with the protein binding site or could be due to the same non-specific binding to micelles possibly encountered with the other two methods described previously. While the T169A mutant is known to not produce the wild type functional response to LCA,[65] experimental data are not available regarding whether binding of LCA is altered in the mutant. This could help to explain the lack of difference in eluted protein in WT and T169A samples, but does not mitigate the fact that this assay also failed to demonstrate functionally folded purified protein had been obtained.

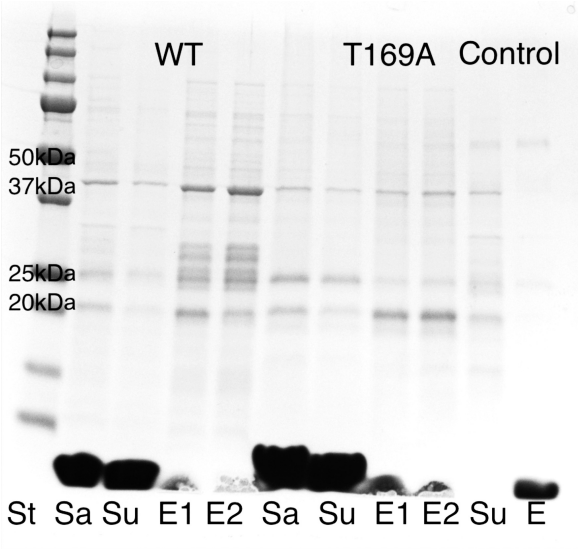


Figure 39. Lithocholic Acid Magnetic Bead β_1 Pull-down
St represents the protein standard. Sa is the sample before exposure to magnetic beads. Su is the sample after exposure to magnetic beads. E1 and E2 are the elution fractions as described in section 0. Control samples were unreacted magnetic beads incubated with protein.

4.4 Conclusions

While expression and purification protocols have been established by this study, a method to test the functionality of the protein expressed in *E. coli* remains elusive. This is, in part, due to the problems associated with all known β_1 ligands being hydrophobic molecules that will readily partition into any membrane or membrane-like environment rather than staying in an aqueous solution[255]. Unfortunately, without a method to adequately assess protein function, structural characterization of the protein with solution NMR was not justified at this time and resources involved in NMR experiments warrant evidence of native-like protein structure before experimentation begins. While these results did not demonstrate a difference between experimental samples and controls, this does not necessarily mean the methods themselves are incapable of assessing protein

function. If a well-dispersed heteronuclear spin quantum correlation (HSQC) NMR spectra can be obtained, experiments titrating known β_1 ligands and observing changes in chemical shift would be sufficient to indicate proper structure and allow further NMR experiments to proceed. The problem could lie in the protein structure obtained from *E. coli*, which would necessitate different expression and purification protocols. Further experimentation with pMCSG19 β_1 protein could prove more successful for functional assessment and structural characterization efforts.

Chapter 5

Discussion, Conclusions, and Recommendations

5.1 Discussion of Findings and Conclusions

5.1.1 Molecular Dynamics Simulations of the BK_{Ca} α Subunit Cytosolic Domain

The current molecular dynamics simulations have provided insight into the importance of specific residues in the stabilization of CLR on the CRAC4 domain in the BK_{Ca} α subunit cytosolic domain. Using simulated constructs that mimic experimental counterparts, variation in the behavior of CLR can be seen while interacting with CRAC4 residues of CTD, trCTD, and K453A and Y450F mutant constructs over 200 ns of simulation. With experimental results showing the trCTD construct is more sensitive than the full WT protein, the simulations provide evidence that the increased stability of CLR on CRAC4 in trCTD protein is caused, in part, by increased hydrogen bonding between K453 and CLR. Loss of this hydrogen bonding partner in K453A simulations resulted in a marked decrease in time CLR spent on CRAC4 residues 450 to 453 in triplicate simulations further supporting the idea that K453 is vital to the CLR sensitivity of the channel and explaining experimental data. However, two events in the simulations remain unexplained. First, the difference in the behavior of cholesterol between the trCTD and Y450F simulations, despite the presence of a hydrogen bonding partner, could be observed in simulation trajectories, but could not be explained quantitatively. Second, the migration of CLR to CRAC6 and CRAC7 in CTD simulations 1 and 2 remain unexplained by this study, although these two CRAC domains are both exposed and closer to CRAC4 than other

CTD CRAC motifs in the BK_{Ca} CTD model and crystal structure. While both of these likely have explanations in their respective simulation trajectories, limits on currently available analysis tools have prevented their description and quantification.

5.1.2 Pharmacophore Modeling of the BK_{Ca} β_1 Subunit

By utilizing the knowledge gained from multiple generations of pharmacophore modeling and improvements in modeling software, there is now a robust tool that can be used to mine conformational databases to find new activators of BK_{Ca} channel activity. The 3rd generation model used in virtual screening has brought us closer to discovering therapeutic lead compounds that can be used to develop novel activators of BK_{Ca} through direct interactions with the β_1 subunit. With tissue specific expression, targeting of the β_1 subunit could provide new means of treating and managing diseases with a potentially smaller side effect profile than existing therapeutics. Additionally, using *in vitro* methods to test compounds discovered by virtual screening with these models will help to produce a new model that can strike a balance between sensitivity and specificity to provide significant enrichment of virtual screening results without discarding too many potentially active compounds. Since the current approach uses a model with high specificity, but low sensitivity, many compounds are being discarded in the initial search. A more balanced model would improve upon the search capabilities of the 3rd generation, as well as the combined 3rd and 6th generation model approach, by not discarding as many actives in the initial search. Furthermore, a more structurally diverse set of compounds that can be used to train or evaluate new models will provide a deeper understanding of the

cholane-steroid interaction site of the β_1 subunit until high-resolution structural data are available.

5.1.3 Expression and Purification of the BK_{Ca} β_1 Subunit in *E. coli*

With no high-resolution structural data available for any of the β subunits of the BK_{Ca} channel, developing and refining expression and purification protocols overcomes a major hurdle in the effort to obtain structural data. Successful expression of the β_1 subunit in *E. coli* was accomplished using optimization of the gene for the host organism and purification from insoluble cell mass was accomplished using IMAC purification techniques and on-column detergent exchange yielding highly pure protein samples stabilized in detergent micelles. Expression of the protein in *E. coli* provides the most cost-effective means to produce quantities of protein sufficient for NMR studies. However, before NMR structural studies can be carried out, proof of native-like activity must be obtained to justify the time and expense of NMR experiments. The methods described in this dissertation failed to show the reversible binding of known β_1 subunit ligands, so structural studies are not yet justified. Since these experiments were not an exhaustive list of available options, the possibility remains that a method to demonstrate protein functionality is within reach and can provide the data necessary to commence structural studies via NMR.

5.2 Recommendations for Future Research

5.2.1 Understanding the Interactions Between Cholesterol and CRAC Motif Residues

Although conclusions can be made about the importance of K453 in the BK_{Ca} CRAC4 motif, an explanation of the role for Y450 in cholesterol recognition

and stabilization remains unresolved. Further studies of the binding energies in this and other CRAC motifs using quantum mechanical methods may provide insight into the role each residue or the combination of residues plays in CLR recognition, but may be limited due to the computational time required to calculate larger atomic systems. Additionally, further simulations of a tetrameric assembly of CTD protein may provide results different from the current study due to the proximity of CRAC4 to the subunit interface in the tetrameric structure. With GPU acceleration of molecular dynamics simulations, a macromolecular simulation of this size is now feasible. Including a membrane-embedded CLR may also provide insight into the mechanism that causes CLR to leave the hydrophobic membrane environment to interact with cytosolic protein domains in BK_{Ca} α subunits, but these simulations may require an extensive time scale to produce an observable effect. More tractable simulations going forward could also include simulation of CLR starting on other CRAC motifs in the CTD protein structure to determine if the behavior of CLR observed on CRAC4 is unique. A few or all of these simulations would significantly further the knowledge of the interaction between CLR and BK_{Ca}.

5.2.2 Improvements to Pharmacophore Models

Armed with newly acquired compounds that will be confirmed to be either active or inactive using *in vitro* electrophysiology studies, improvements to the existing models or even entirely new models could be developed to improve upon the current understanding of the β_1 subunit cholane-steroid recognition site pharmacophore. Increased structural diversity in compounds used to train models will provide new representations of the pharmacophore and hopefully

improve performance and produce a model that is more balanced between sensitivity and specificity and will not discard as many active compounds. With current models being overly discriminating or overly lax, a new model based on the most current and most active compounds could provide this more balanced approach. Furthermore, use of newer pharmacophore development tools, such as the ROCS software from OpenEye Scientific, could improve virtual screening methods by factoring molecular shape and volume into models along with ligand features similar to those used in MOE and discussed previously. Another software package of interest is the academically developed HPCC software,[258] which has proven as efficient as the commercially developed ROCS software in benchmarking tests. Using these new approaches, the discovery of therapeutic lead compounds to target the β_1 subunit cholane-steroid recognition site would be accelerated and could lead to clinically relevant therapeutics in the long term.

5.2.3 Functional Studies and Structural Characterization of the BK_{Ca} β_1 Subunit

Since the functional studies discussed in this dissertation did not show reversible binding of ligands, a new approach is needed to move forward with structural characterization. In many cases, a well-dispersed heteronuclear single quantum coherence (HSQC) spectrum is sufficient to indicate the existence of secondary structure in protein NMR, but this could be a significant challenge due to the helical nature of the β_1 subunit. One option to overcome problems associated with overlapped chemical shifts in helical proteins would be to titrate a known β_1 ligand into a protein sample in increasing concentrations to measure changes in chemical shift in an HSQC spectra. Due to the hydrophobic nature of

the known β_1 ligands and the presence of detergent micelles, functional assays have proven fruitless, but this method should produce changes in chemical shifts only in residues associated with binding in the cholane-steroid recognition site. Thus non-specific interactions which would have more global impact on protein chemical shifts could be differentiated from specific interactions without the need to displace a hydrophobic ligand from both the protein surface and the micelle.

If well-ordered secondary structure and ligand binding are not observed in HSQC experiments, varying the purification protocol to include more mild detergents may produce higher quality, but less pure protein. Some more recent methods include gentle sonication of insoluble cell mass to quickly solubilize proteins without the use of denaturing conditions or harsh detergents. Since SDS interferes with IMAC binding, changing the solubilization detergent would probably yield a less pure sample that would have to undergo a secondary purification step, for which a protocol would have to be developed. However, if less harsh detergent conditions produce a higher quality, properly folded sample that can lead to structural studies, any loss of protein in secondary purification would be justifiable to eventually produce high-resolution structural data.

Recent advancements in the structural characterization of membrane proteins by solution NMR should provide high quality data that will lead to protein structure. An alternative would be the use of solid state NMR of proteins imbedded in membrane patches if solution NMR techniques fail. For an extensive discussion of membrane mimetics for solution NMR see Warschawski *et al*[259] and for a review of integral membrane proteins studied using NMR see Kang *et*

al.[222] With the refined protocol for expression and purification of β_1 protein in *E. coli* described in this dissertation, obtaining high resolution structural data for the subunit, which would provide invaluable insight into structure and also assist in therapeutic lead compound development, is one step closer to completion.

References

(1)McManus, O. B. *Journal of bioenergetics and biomembranes* **1991**, *23*, 537-560.

(2)Knaus, H. G.; Schwarzer, C.; Koch, R. O.; Eberhart, A.; Kaczorowski, G. J.; Glossmann, H.; Wunder, F.; Pongs, O.; Garcia, M. L.; Sperk, G. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **1996**, *16*, 955-963.

(3)Eichhorn, B.; Dobrev, D. *Naunyn-Schmiedeberg's archives of pharmacology* **2007**, *376*, 145-155.

(4)Marty, A. *Nature* **1981**, *291*, 497-500.

(5)Gardos, G. *Biochimica et biophysica acta* **1958**, *30*, 653-654.

(6)Ishii, T. M.; Silvia, C.; Hirschberg, B.; Bond, C. T.; Adelman, J. P.; Maylie, J. *Proceedings of the National Academy of Sciences of the United States of America* **1997**, *94*, 11651-11656.

(7)Blatz, A. L.; Magleby, K. L. *Nature* **1986**, *323*, 718-720.

(8)Meech, R. W.; Strumwas.F *Fed Proc* **1970**, *29*, A834-&.

(9)Knaus, H. G.; Eberhart, A.; Koch, R. O.; Munujos, P.; Schmalhofer, W. A.; Warmke, J. W.; Kaczorowski, G. J.; Garcia, M. L. *The Journal of biological chemistry* **1995**, *270*, 22434-22439.

(10)Papazian, D. M.; Schwarz, T. L.; Tempel, B. L.; Jan, Y. N.; Jan, L. Y. *Science* **1987**, *237*, 749-753.

(11)Pongs, O.; Kecskemethy, N.; Muller, R.; Krah-Jentgens, I.; Baumann, A.; Kiltz, H. H.; Canal, I.; Llamazares, S.; Ferrus, A. *The EMBO journal* **1988**, *7*, 1087-1096.

(12)Meera, P.; Wallner, M.; Song, M.; Toro, L. *Proceedings of the National Academy of Sciences of the United States of America* **1997**, *94*, 14066-14071.

(13)Ma, Z.; Lou, X. J.; Horrigan, F. T. *J Gen Physiol* **2006**, *127*, 309-328.

(14)Wallner, M.; Meera, P.; Toro, L. *Proceedings of the National Academy of Sciences of the United States of America* **1996**, *93*, 14922-14927.

(15)Morrow, J. P.; Zakharov, S. I.; Liu, G.; Yang, L.; Sok, A. J.; Marx, S. O. *Proceedings of the National Academy of Sciences of the United States of America* **2006**, *103*, 5096-5101.

(16)Koval, O. M.; Fan, Y.; Rothberg, B. S. *J Gen Physiol* **2007**, *129*, 209-220.

(17)Yuan, P.; Leonetti, M. D.; Pico, A. R.; Hsiung, Y.; MacKinnon, R. *Science* **2010**, *329*, 182-186.

(18)Jiang, Y.; Pico, A.; Cadene, M.; Chait, B. T.; MacKinnon, R. *Neuron* **2001**, *29*, 593-601.

(19)Xia, X. M.; Zeng, X.; Lingle, C. J. *Nature* **2002**, *418*, 880-884.

(20)Schreiber, M.; Salkoff, L. *Biophysical journal* **1997**, *73*, 1355-1363.

(21)Brayden, J. E.; Nelson, M. T. *Science* **1992**, *256*, 532-535.

(22)Perez, G. J.; Bonev, A. D.; Patlak, J. B.; Nelson, M. T. *J Gen Physiol* **1999**, *113*, 229-238.

(23)Jaggar, J. H.; Wellman, G. C.; Heppner, T. J.; Porter, V. A.; Perez, G. J.; Gollasch, M.; Kleppisch, T.; Rubart, M.; Stevenson, A. S.; Lederer, W. J.; Knot, H. J.; Bonev, A. D.; Nelson, M. T. *Acta physiologica Scandinavica* **1998**, *164*, 577-587.

(24)Knot, H. J.; Standen, N. B.; Nelson, M. T. *The Journal of physiology* **1998**, *508 (Pt 1)*, 211-221.

(25)Girouard, H.; Bonev, A. D.; Hannah, R. M.; Meredith, A.; Aldrich, R. W.; Nelson, M. T. *Proceedings of the National Academy of Sciences of the United States of America* **2010**, *107*, 3811-3816.

(26)Filosa, J. A.; Bonev, A. D.; Straub, S. V.; Meredith, A. L.; Wilkerson, M. K.; Aldrich, R. W.; Nelson, M. T. *Nature neuroscience* **2006**, *9*, 1397-1403.

(27)Navaratnam, D. S.; Bell, T. J.; Tu, T. D.; Cohen, E. L.; Oberholtzer, J. C. *Neuron* **1997**, *19*, 1077-1085.

(28)Rosenblatt, K. P.; Sun, Z. P.; Heller, S.; Hudspeth, A. J. *Neuron* **1997**, *19*, 1061-1075.

(29)Tseng-Crank, J.; Foster, C. D.; Krause, J. D.; Mertz, R.; Godinot, N.; DiChiara, T. J.; Reinhart, P. H. *Neuron* **1994**, *13*, 1315-1330.

(30)Orio, P.; Rojas, P.; Ferreira, G.; Latorre, R. *News in physiological sciences : an international journal of physiology produced jointly by the International Union of Physiological Sciences and the American Physiological Society* **2002**, *17*, 156-161.

(31)Brenner, R.; Jegla, T. J.; Wickenden, A.; Liu, Y.; Aldrich, R. W. *The Journal of biological chemistry* **2000**, *275*, 6453-6461.

(32)Tseng-Crank, J.; Godinot, N.; Johansen, T. E.; Ahring, P. K.; Strobaek, D.; Mertz, R.; Foster, C. D.; Olesen, S. P.; Reinhart, P. H. *Proceedings of the National Academy of Sciences of the United States of America* **1996**, *93*, 9200-9205.

(33)Gonzalez, C.; Baez-Nieto, D.; Valencia, I.; Oyarzun, I.; Rojas, P.; Naranjo, D.; Latorre, R. *Comprehensive Physiology* **2012**, *2*, 2087-2149.

(34)Ding, J. P.; Li, Z. W.; Lingle, C. J. *Biophysical journal* **1998**, *74*, 268-289.

(35)Liu, G.; Niu, X.; Wu, R. S.; Chudasama, N.; Yao, Y.; Jin, X.; Weinberg, R.; Zakharov, S. I.; Motoike, H.; Marx, S. O.; Karlin, A. *J Gen Physiol* **2010**, *135*, 449-459.

(36)Sun, X. H.; Shi, J. Y.; Delaloye, K.; Yang, X.; Yang, H. H.; Zhang, G. H.; Cui, J. M. *Journal of Neuroscience* **2013**, *33*, 11253-11261.

(37)Orio, P.; Torres, Y.; Rojas, P.; Carvacho, I.; Garcia, M. L.; Toro, L.; Valverde, M. A.; Latorre, R. *J Gen Physiol* **2006**, *127*, 191-204.

(38)Orio, P.; Latorre, R. *J Gen Physiol* **2005**, *125*, 395-411.

(39)McManus, O. B.; Helms, L. M.; Pallanck, L.; Ganetzky, B.; Swanson, R.; Leonard, R. J. *Neuron* **1995**, *14*, 645-650.

(40)Jiang, Z.; Wallner, M.; Meera, P.; Toro, L. *Genomics* **1999**, *55*, 57-67.

(41)Brenner, R.; Perez, G. J.; Bonev, A. D.; Eckman, D. M.; Kosek, J. C.; Wiler, S. W.; Patterson, A. J.; Nelson, M. T.; Aldrich, R. W. *Nature* **2000**, *407*, 870-876.

(42)Pluznick, J. L.; Wei, P.; Carmines, P. K.; Sansom, S. C. *American journal of physiology. Renal physiology* **2003**, *284*, F1274-1279.

(43)Lee, U. S.; Shi, J.; Cui, J. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **2010**, *30*, 16170-16179.

(44)Wallner, M.; Meera, P.; Toro, L. *Proceedings of the National Academy of Sciences of the United States of America* **1999**, *96*, 4137-4142.

(45)Yang, H.; Zhang, G.; Shi, J.; Lee, U. S.; Delaloye, K.; Cui, J. *Biophysical journal* **2008**, *94*, 4678-4687.

(46)Lingle, C. J.; Zeng, X. H.; Ding, J. P.; Xia, X. M. *J Gen Physiol* **2001**, *117*, 583-606.

(47)Xia, X. M.; Ding, J. P.; Lingle, C. J. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **1999**, *19*, 5255-5264.

(48)Uebele, V. N.; Lagrutta, A.; Wade, T.; Figueroa, D. J.; Liu, Y.; McKenna, E.; Austin, C. P.; Bennett, P. B.; Swanson, R. *The Journal of biological chemistry* **2000**, *275*, 23211-23218.

(49)Behrens, R.; Nolting, A.; Reimann, F.; Schwarz, M.; Waldschutz, R.; Pongs, O. *FEBS letters* **2000**, *474*, 99-106.

(50)Riazi, M. A.; Brinkman-Mills, P.; Johnson, A.; Naylor, S. L.; Minoshima, S.; Shimizu, N.; Baldini, A.; McDermid, H. E. *Genomics* **1999**, *62*, 90-94.

(51)Yan, J.; Aldrich, R. W. *Nature* **2010**, *466*, 513-516.

(52)Yan, J.; Aldrich, R. W. *Proceedings of the National Academy of Sciences of the United States of America* **2012**, *109*, 7917-7922.

(53)Wu, Y.; Yang, Y.; Ye, S.; Jiang, Y. *Nature* **2010**, *466*, 393-397.

(54)Shi, P.; Li, D.; Lai, C.; Zhang, L.; Tian, C. *Biochemical and biophysical research communications* **2013**, *437*, 408-412.

(55)Wang, L.; Sigworth, F. J. *Nature* **2009**, *461*, 292-295.

(56)Bentrop, D.; Beyermann, M.; Wissmann, R.; Fakler, B. *The Journal of biological chemistry* **2001**, *276*, 42116-42121.

(57)Contreras, G. F.; Castillo, K.; Enrique, N.; Carrasquel-Ursulaez, W.; Castillo, J. P.; Milesi, V.; Neely, A.; Alvarez, O.; Ferreira, G.; Gonzalez, C.; Latorre, R. *Channels* **2013**, *7*, 442-458.

(58)Dopico, A. M.; Bukiya, A. N.; Singh, A. K. *Pharmacology & therapeutics* **2012**, *135*, 133-150.

(59)Cox, D. H.; Cui, J.; Aldrich, R. W. *J Gen Physiol* **1997**, *110*, 257-281.

(60)Meera, P.; Wallner, M.; Jiang, Z.; Toro, L. *FEBS letters* **1996**, *385*, 127-128.

(61)Wallner, M.; Meera, P.; Ottolia, M.; Kaczorowski, G. J.; Latorre, R.; Garcia, M. L.; Stefani, E.; Toro, L. *Receptors & channels* **1995**, *3*, 185-199.

(62)Dworetzky, S. I.; Boissard, C. G.; Lum-Ragan, J. T.; McKay, M. C.; Post-Munson, D. J.; Trojnacki, J. T.; Chang, C. P.; Gribkoff, V. K. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **1996**, *16*, 4543-4550.

(63)Gruslova, A.; Semenov, I.; Wang, B. *J Gen Physiol* **2012**, *139*, 57-67.

(64)Qian, X.; Nimigean, C. M.; Niu, X.; Moss, B. L.; Magleby, K. L. *J Gen Physiol* **2002**, *120*, 829-843.

(65)Bukiya, A. N.; Singh, A. K.; Parrill, A. L.; Dopico, A. M. *Proceedings of the National Academy of Sciences of the United States of America* **2011**, *108*, 20207-20212.

(66)Latorre, R.; Brauchi, S. *Biological research* **2006**, *39*, 385-401.

(67)Braun, M.; Ramracheya, R.; Bengtsson, M.; Zhang, Q.; Karanauskaite, J.; Partridge, C.; Johnson, P. R.; Rorsman, P. *Diabetes* **2008**, *57*, 1618-1628.

(68)Swensen, A. M.; Bean, B. P. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **2003**, *23*, 9650-9663.

(69)Womack, M. D.; Khodakhah, K. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **2004**, *24*, 3511-3521.

(70)Pinterova, M.; Behuliak, M.; Kunes, J.; Zicha, J. *Physiological research / Academia Scientiarum Bohemoslovaca* **2014**.

(71)Ramanathan, K.; Michael, T. H.; Fuchs, P. A. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **2000**, *20*, 1675-1684.

(72)Williams, S. E.; Wootton, P.; Mason, H. S.; Bould, J.; Iles, D. E.; Riccardi, D.; Peers, C.; Kemp, P. J. *Science* **2004**, *306*, 2093-2097.

(73)Samengo, I.; Curro, D.; Barrese, V.; Taglialatela, M.; Martire, M. *Neurochemical research* **2014**.

(74)Mannowetz, N.; Naidoo, N. M.; Choo, S. A.; Smith, J. F.; Lishko, P. V. *eLife* **2013**, *2*, e01009.

(75)Liang, F.; Schulte, B. A.; Qu, C.; Hu, W.; Shen, Z. *Neuroscience* **2005**, *135*, 263-271.

(76)Tricarico, D.; Mele, A.; Calzolaro, S.; Cannone, G.; Camerino, G. M.; Dinardo, M. M.; Latorre, R.; Conte Camerino, D. *Plos One* **2013**, *8*, e69551.

(77)Woda, C. B.; Bragin, A.; Kleyman, T. R.; Satlin, L. M. *American journal of physiology. Renal physiology* **2001**, *280*, F786-793.

(78)Matzkin, M. E.; Lauf, S.; Spinnler, K.; Rossi, S. P.; Kohn, F. M.; Kunz, L.; Calandra, R. S.; Frungieri, M. B.; Mayerhofer, A. *Molecular and cellular endocrinology* **2013**, *367*, 41-49.

(79)Zhang, L.; Li, X.; Zhou, R.; Xing, G. *Medical hypotheses* **2006**, *67*, 41-43.

(80)Laumonnier, F.; Roger, S.; Guerin, P.; Molinari, F.; M'Rad, R.; Cahard, D.; Belhadj, A.; Halayem, M.; Persico, A. M.; Elia, M.; Romano, V.; Holbert, S.;

Andres, C.; Chaabouni, H.; Colleaux, L.; Constant, J.; Le Guennec, J. Y.; Briault, S. *The American journal of psychiatry* **2006**, *163*, 1622-1629.

(81)Lorenz, S.; Heils, A.; Kasper, J. M.; Sander, T. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics* **2007**, *144B*, 10-13.

(82)Du, W.; Bautista, J. F.; Yang, H.; Diez-Sampedro, A.; You, S. A.; Wang, L.; Kotagal, P.; Luders, H. O.; Shi, J.; Cui, J.; Richerson, G. B.; Wang, Q. K. *Nature genetics* **2005**, *37*, 733-738.

(83)Diez-Sampedro, A.; Silverman, W. R.; Bautista, J. F.; Richerson, G. B. *J Neurophysiol* **2006**, *96*, 1507-1516.

(84)Lee, U. S.; Cui, J. *The Journal of physiology* **2009**, *587*, 1481-1498.

(85)Wang, B.; Rothberg, B. S.; Brenner, R. *Journal of General Physiology* **2009**, *133*, 283-294.

(86)Yang, J.; Krishnamoorthy, G.; Saxena, A.; Zhang, G.; Shi, J.; Yang, H.; Delaloye, K.; Sept, D.; Cui, J. *Neuron* **2010**, *66*, 871-883.

(87)Kronengold, J.; Kaczmarek, L. K. *Neuron* **2010**, *66*, 817-818.

(88)Brenner, R.; Chen, Q. H.; Vilaythong, A.; Toney, G. M.; Noebels, J. L.; Aldrich, R. W. *Nature neuroscience* **2005**, *8*, 1752-1759.

(89)Pluger, S.; Faulhaber, J.; Furstenau, M.; Lohn, M.; Waldschutz, R.; Gollasch, M.; Haller, H.; Luft, F. C.; Ehmke, H.; Pongs, O. *Circulation research* **2000**, *87*, E53-60.

(90)Amberg, G. C.; Santana, L. F. *Circulation research* **2003**, *93*, 965-971.

(91)Amberg, G. C.; Bonev, A. D.; Rossow, C. F.; Nelson, M. T.; Santana, L. F. *The Journal of clinical investigation* **2003**, *112*, 717-724.

(92)Grimm, P. R.; Irsik, D. L.; Settles, D. C.; Holtzclaw, J. D.; Sansom, S. C. *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106*, 11800-11805.

(93)Fernandez-Fernandez, J. M.; Tomas, M.; Vazquez, E.; Orio, P.; Latorre, R.; Senti, M.; Marrugat, J.; Valverde, M. A. *The Journal of clinical investigation* **2004**, *113*, 1032-1039.

(94)Ledoux, J.; Werner, M. E.; Brayden, J. E.; Nelson, M. T. *Physiology* **2006**, *21*, 69-78.

(95)Werner, M. E.; Zvara, P.; Meredith, A. L.; Aldrich, R. W.; Nelson, M. T. *The Journal of physiology* **2005**, *567*, 545-556.

(96)Werner, M. E.; Meredith, A. L.; Aldrich, R. W.; Nelson, M. T. *American journal of physiology. Regulatory, integrative and comparative physiology* **2008**, *295*, R181-188.

(97)Meredith, A. L.; Thorneloe, K. S.; Werner, M. E.; Nelson, M. T.; Aldrich, R. W. *The Journal of biological chemistry* **2004**, *279*, 36746-36752.

(98)Thorneloe, K. S.; Meredith, A. L.; Knorn, A. M.; Aldrich, R. W.; Nelson, M. T. *American journal of physiology. Renal physiology* **2005**, *289*, F604-610.

(99)Zakoji, H. K., H.; Yoshiyama, M.; Takeda, M.; Araki, I. *Open J Urology* **2013**, *3*, 47-52.

(100)Semenov, I.; Wang, B.; Herlihy, J. T.; Brenner, R. *American journal of physiology. Lung cellular and molecular physiology* **2006**, *291*, L802-810.

(101)Semenov, I.; Wang, B.; Herlihy, J. T.; Brenner, R. *The Journal of physiology* **2011**, *589*, 1803-1817.

(102)Vang, A.; Mazer, J.; Casserly, B.; Choudhary, G. *Vascular pharmacology* **2010**, *53*, 122-129.

(103)Grimm, P. R.; Foutz, R. M.; Brenner, R.; Sansom, S. C. *American journal of physiology. Renal physiology* **2007**, *293*, F350-359.

(104)Wen, D.; Cornelius, R. J.; Yuan, Y.; Sansom, S. C. *American journal of physiology. Renal physiology* **2013**, *305*, F463-476.

(105)Holtzclaw, J. D.; Grimm, P. R.; Sansom, S. C. *Journal of the American Society of Nephrology : JASN* **2010**, *21*, 634-645.

(106)Pluznick, J. L.; Sansom, S. C. *American journal of physiology. Renal physiology* **2006**, *291*, F517-529.

(107)Kryshtal, D. A.; Paduraru, O. M.; Boldyriev, O. I.; Kit, O.; Rekalov, V. V.; Shuba Ia, M. *Fiziolohichnyi zhurnal* **2011**, *57*, 25-32.

(108)Yi, F.; Wang, H.; Chai, Q.; Wang, X.; Shen, W. K.; Willis, M. S.; Lee, H. C.; Lu, T. *The Journal of biological chemistry* **2014**.

(109)Foutz, R. M.; Grimm, P. R.; Sansom, S. C. *American journal of physiology. Renal physiology* **2008**, *294*, F1465-1472.

(110)Burnham, M. P.; Johnson, I. T.; Weston, A. H. *American journal of physiology. Heart and circulatory physiology* **2006**, *290*, H1520-1527.

(111)Lu, T.; Wang, X. L.; He, T.; Zhou, W.; Kaduce, T. L.; Katusic, Z. S.; Spector, A. A.; Lee, H. C. *Diabetes* **2005**, *54*, 2155-2163.

(112)Howitt, L.; Sandow, S. L.; Grayson, T. H.; Ellis, Z. E.; Morris, M. J.; Murphy, T. V. *American journal of physiology. Heart and circulatory physiology* **2011**, *301*, H29-40.

(113)Miller, C.; Moczydlowski, E.; Latorre, R.; Phillips, M. *Nature* **1985**, *313*, 316-318.

(114)Candia, S.; Garcia, M. L.; Latorre, R. *Biophysical journal* **1992**, *63*, 583-590.

(115)Bontems, F.; Gilquin, B.; Roumestand, C.; Menez, A.; Toma, F. *Biochemistry* **1992**, *31*, 7756-7764.

(116)Meera, P.; Wallner, M.; Toro, L. *Proceedings of the National Academy of Sciences of the United States of America* **2000**, *97*, 5562-5567.

(117)Gan, G.; Yi, H.; Chen, M.; Sun, L.; Li, W.; Wu, Y.; Ding, J. *The Journal of biological chemistry* **2008**, *283*, 24177-24184.

(118)Hanner, M.; Schmalhofer, W. A.; Munujos, P.; Knaus, H. G.; Kaczorowski, G. J.; Garcia, M. L. *Proceedings of the National Academy of Sciences of the United States of America* **1997**, *94*, 2853-2858.

(119)Hanner, M.; Vianna-Jorge, R.; Kamassah, A.; Schmalhofer, W. A.; Knaus, H. G.; Kaczorowski, G. J.; Garcia, M. L. *The Journal of biological chemistry* **1998**, *273*, 16289-16296.

(120)Knaus, H. G.; McManus, O. B.; Lee, S. H.; Schmalhofer, W. A.; Garcia-Calvo, M.; Helms, L. M.; Sanchez, M.; Giangiacomo, K.; Reuben, J. P.; Smith, A. B., 3rd; et al. *Biochemistry* **1994**, *33*, 5819-5828.

(121)Olesen, S. P.; Munch, E.; Moldt, P.; Drejer, J. *Eur J Pharmacol* **1994**, *251*, 53-59.

(122)Lee, K.; Rowe, I. C.; Ashford, M. L. *Eur J Pharmacol* **1995**, *280*, 215-219.

(123)Siemer, C.; Bushfield, M.; Newgreen, D.; Grissmer, S. *The Journal of membrane biology* **2000**, *173*, 57-66.

(124)Hu, S.; Kim, H. S.; Fink, C. A. *Eur J Pharmacol* **1995**, *294*, 357-360.

(125)Hu, S.; Kim, H. S. *Eur J Pharmacol* **1996**, *318*, 461-468.

(126)Gessner, G.; Cui, Y. M.; Otani, Y.; Ohwada, T.; Soom, M.; Hoshi, T.; Heinemann, S. H. *Proceedings of the National Academy of Sciences of the United States of America* **2012**, *109*, 3552-3557.

(127)Revermann, M.; Neofitidou, S.; Kirschning, T.; Schloss, M.; Brandes, R. P.; Hofstetter, C. *Plos One* **2014**, *9*, e86636.

(128)Jensen, B. S. *CNS drug reviews* **2002**, *8*, 353-360.

(129)Cheney, J. A.; Weisser, J. D.; Bareyre, F. M.; Laurer, H. L.; Saatman, K. E.; Raghupathi, R.; Gribkoff, V.; Starrett, J. E., Jr.; McIntosh, T. K. *Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism* **2001**, *21*, 396-403.

(130)Dopico, A. M.; Walsh, J. V., Jr.; Singer, J. J. *J Gen Physiol* **2002**, *119*, 251-273.

(131)Bukiya, A. N.; Liu, J.; Toro, L.; Dopico, A. M. *Mol Pharmacol* **2007**, *72*, 359-369.

(132)Bukiya, A. N.; Vaithianathan, T.; Toro, L.; Dopico, A. M. *FEBS letters* **2008**, *582*, 673-678.

(133)Bukiya, A. N.; Vaithianathan, T.; Toro, L.; Dopico, A. M. *Biochemical and biophysical research communications* **2009**, *390*, 995-1000.

(134)Bukiya, A. N.; McMillan, J.; Parrill, A. L.; Dopico, A. M. *J Lipid Res* **2008**, *49*, 2441-2451.

(135)Chiang, J. Y. *Comprehensive Physiology* **2013**, *3*, 1191-1212.

(136)Bolotina, V.; Omelyanenko, V.; Heyes, B.; Ryan, U.; Bregestovski, P. *Pflugers Archiv : European journal of physiology* **1989**, *415*, 262-268.

(137)Singh, A. K.; McMillan, J.; Bukiya, A. N.; Burton, B.; Parrill, A. L.; Dopico, A. M. *The Journal of biological chemistry* **2012**.

(138)Levitan, I.; Fang, Y.; Rosenhouse-Dantsker, A.; Romanenko, V. In *Cholesterol Binding and Cholesterol Transport Proteins:*; Harris, J. R., Ed.; Springer Netherlands: 2010; Vol. 51, p 509-549.

(139)McManus, O. B.; Harris, G. H.; Giangiacomo, K. M.; Feigenbaum, P.; Reuben, J. P.; Addy, M. E.; Burka, J. F.; Kaczorowski, G. J.; Garcia, M. L. *Biochemistry* **1993**, *32*, 6128-6133.

(140)Giangiacomo, K. M.; Kamassah, A.; Harris, G.; McManus, O. B. *Journal of General Physiology* **1998**, *112*, 485-501.

(141)Bukiya, A. N.; McMillan, J. E.; Fedinec, A. L.; Patil, S. A.; Miller, D. D.; Leffler, C. W.; Parrill, A. L.; Dopico, A. M. *Mol Pharmacol* **2013**, *83*, 1030-1044.

(142)Singh, S. B.; Goetz, M. A.; Zink, D. L.; Dombrowski, A. W.; Polishook, J. D.; Garcia, M. L.; Schmalhofer, W.; Mcmanus, O. B.; Kaczorowski, G. J. *J Chem Soc Perk T 1* **1994**, 3349-3352.

(143)Valverde, M. A.; Rojas, P.; Amigo, J.; Cosmelli, D.; Orio, P.; Bahamonde, M. I.; Mann, G. E.; Vergara, C.; Latorre, R. *Science* **1999**, *285*, 1929-1931.

(144)Azenabor, A. A.; Chaudhry, A. U. *Journal of reproductive immunology* **2003**, *59*, 17-28.

(145)Dick, G. M.; Rossow, C. F.; Smirnov, S.; Horowitz, B.; Sanders, K. M. *The Journal of biological chemistry* **2001**, *276*, 34594-34599.

(146)Dick, G. M.; Sanders, K. M. *The Journal of biological chemistry* **2001**, *276*, 44835-44840.

(147)Korovkina, V. P.; Brainard, A. M.; Ismail, P.; Schmidt, T. J.; England, S. K. *The Journal of biological chemistry* **2004**, *279*, 1217-1223.

(148)Dick, G. M. *British journal of pharmacology* **2002**, *136*, 961-964.

(149)Rottgen, T. S.; Fancher, I. S.; Asano, S.; Widlanski, T. S.; Dick, G. M. *Channels* **2014**, *8*.

(150)Nardi, A.; Olesen, S. P. *Current medicinal chemistry* **2008**, *15*, 1126-1146.

(151)Stein, D. L. *Proceedings of the National Academy of Sciences of the United States of America* **1985**, *82*, 3670-3672.

(152)Frauenfelder, H.; Chen, G.; Berendzen, J.; Fenimore, P. W.; Jansson, H.; McMahon, B. H.; Stroe, I. R.; Swenson, J.; Young, R. D. *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106*, 5129-5134.

(153)Kern, D.; Zuiderweg, E. R. *Current opinion in structural biology* **2003**, *13*, 748-757.

(154)Abbruzzetti, S.; Faggiano, S.; Bruno, S.; Spyrakis, F.; Mozzarelli, A.; Dewilde, S.; Moens, L.; Viappiani, C. *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106*, 18984-18989.

(155)Bidon-Chanal, A.; Marti, M. A.; Estrin, D. A.; Luque, F. J. *J Am Chem Soc* **2007**, *129*, 6782-6788.

(156)Marti, M. A.; Gonzalez Lebrero, M. C.; Roitberg, A. E.; Estrin, D. A. *J Am Chem Soc* **2008**, *130*, 1611-1618.

(157)Morales-Quintana, L.; Nunez-Tobar, M. X.; Moya-Leon, M. A.; Herrera, R. *J Chem Inf Model* **2013**, *53*, 2689-2700.

(158)Goh, B. C.; Rynkiewicz, M. J.; Cafarella, T. R.; White, M. R.; Hartshorn, K. L.; Allen, K.; Crouch, E. C.; Calin, O.; Seeberger, P. H.; Schulten, K.; Seaton, B. A. *Biochemistry* **2013**, *52*, 8527-8538.

(159)Alder, B. J.; Wainwright, T. E. *The Journal of Chemical Physics* **1959**, *31*, 459-466.

(160)Newton, I. *Philosophiae Naturalis Principia Mathematica* London, 1686.

(161)Sutmann, G. In *Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms*; Grotendorst, J. M., D.; Muramatsu, A., Ed.; John von Neumann Institute for Computing: Julich, 2002; Vol. 10, p 211-254.

(162)Case, D. A.; Cheatham, T. E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J Comput Chem* **2005**, *26*, 1668-1688.

(163)Jones, J. E. *Proceedings of the Royal Society of London. Series A* **1924**, *106*, 463-477.

(164)Darden, T.; York, D.; Pedersen, L. *J Chem Phys* **1993**, *98*, 10089-10092.

(165)Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712-725.

(166)Patra, M.; Karttunen, M.; Hyvonen, M. T.; Falck, E.; Lindqvist, P.; Vattulainen, I. *Biophysical journal* **2003**, *84*, 3636-3645.

(167)Wermuth, G.; Ganellin, C. R.; Lindberg, P.; Mitscher, L. A. *Pure Appl Chem* **1998**, *70*, 1129-1143.

(168)Beckett, A. H. *Fortschritte der Arzneimittelforschung. Progress in drug research. Progres des recherches pharmaceutiques* **1959**, *1*, 455-530.

(169)Mayer, D.; Naylor, C. B.; Motoc, I.; Marshall, G. R. *J Comput Aided Mol Des* **1987**, *1*, 3-16.

(170)Guner OF.; Henrey, D. In *Pharmacophore Perception, Development, and Use in Drug Design*; Guner, O., Ed.; International University Line: La Jolla, CA, 2000, p 191-211.

(171)Yang, S. Y. *Drug Discovery Today* **2010**, *15*, 444-450.

(172)Wolber, G.; Seidel, T.; Bendix, F.; Langer, T. *Drug Discov Today* **2008**, *13*, 23-29.

(173)Langer, T. H., R. D. *Pharmacophores and Pharmacophore Searches*; Wiley-VCH: Weinheim, Germany, 2006; Vol. 32.

(174)Wallin, E.; von Heijne, G. *Protein science : a publication of the Protein Society* **1998**, *7*, 1029-1038.

(175)Deisenhofer, J.; Epp, O.; Miki, K.; Huber, R.; Michel, H. *Nature* **1985**, *318*, 618-624.

(176)Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic acids research* **2000**, *28*, 235-242.

(177)Wimley, W. C.; White, S. H. *Biochemistry* **2000**, *39*, 4432-4442.

(178)White, S. H.; von Heijne, G. *Current opinion in structural biology* **2004**, *14*, 397-404.

(179)von Heijne, G. *Quarterly reviews of biophysics* **1999**, *32*, 285-307.

(180)Wagner, S.; Bader, M. L.; Drew, D.; de Gier, J. W. *Trends in biotechnology* **2006**, *24*, 364-371.

(181)Wagner, S.; Baars, L.; Ytterberg, A. J.; Klussmeier, A.; Wagner, C. S.; Nord, O.; Nygren, P. A.; van Wijk, K. J.; de Gier, J. W. *Molecular & cellular proteomics : MCP* **2007**, *6*, 1527-1550.

(182)Wagner, S.; Klepsch, M. M.; Schlegel, S.; Appel, A.; Draheim, R.; Tarry, M.; Hogbom, M.; van Wijk, K. J.; Slotboom, D. J.; Persson, J. O.; de Gier, J. W. *Proceedings of the National Academy of Sciences of the United States of America* **2008**, *105*, 14371-14376.

(183)Thornhill, W. B.; Wu, M. B.; Jiang, X.; Wu, X.; Morgan, P. T.; Margiotta, J. F. *The Journal of biological chemistry* **1996**, *271*, 19093-19098.

(184)Imperiali, B.; O'Connor, S. E. *Curr Opin Chem Biol* **1999**, *3*, 643-649.

(185)Hagen, B. M.; Sanders, K. M. *American journal of physiology. Cell physiology* **2006**, *291*, C750-756.

(186)Epand, R. M. *Progress in lipid research* **2006**, *45*, 279-294.

(187)Singh, A. K.; McMillan, J.; Bukiya, A. N.; Burton, B.; Parrill, A. L.; Dopico, A. M. *The Journal of biological chemistry* **2012**, *287*, 20509-20521.

(188)Yuan, C.; Chen, M.; Covey, D. F.; Johnston, L. J.; Treistman, S. N. *Plos One* **2011**, *6*, e27572.

(189)Bukiya, A. N.; Belani, J. D.; Rychnovsky, S.; Dopico, A. M. *J Gen Physiol* **2011**, *137*, 93-110.

(190)Apweiler, R.; Bateman, A.; Martin, M. J.; O'Donovan, C.; Magrane, M.; Alam-Faruque, Y.; Alpi, E.; Antunes, R.; Arganiska, J.; Casanova, E. B.; Bely, B.; Bingley, M.; Bonilla, C.; Britto, R.; Bursteinas, B.; Chan, W. M.; Chavali, G.; Cibrian-Uhalte, E.; Da Silva, A.; De Giorgi, M.; Fazzini, F.; Gane, P.; Castro, L. G.; Garmiri, P.; Hatton-Ellis, E.; Hieta, R.; Huntley, R.; Legge, D.; Liu, W. D.; Luo, J.; MacDougall, A.; Mutowo, P.; Nightingale, A.; Orchard, S.; Pichler, K.; Poggioli, D.; Pundir, S.; Pureza, L.; Qi, G. Y.; Rosanoff, S.; Sawford, T.; Shypitsyna, A.; Turner, E.; Volynkin, V.; Wardell, T.; Watkins, X.; Zellner, H.; Corbett, M.; Donnelly, M.; Van Rensburg, P.; Goujon, M.; McWilliam, H.; Lopez, R.; Xenarios, I.; Bougueleret, L.; Bridge, A.; Poux, S.; Redaschi, N.; Aimo, L.; Auchincloss, A.; Axelsen, K.; Bansal, P.; Baratin, D.; Binz, P. A.; Blatter, M. C.; Boeckmann, B.; Bolleman, J.; Boutet, E.; Breuza, L.; Casal-Casas, C.; de Castro, E.; Cerutti, L.; Coudert, E.; Cuche, B.; Doche, M.; Dornevil, D.; Duvaud, S.; Estreicher, A.; Famiglietti, L.; Feuermann, M.; Gasteiger, E.; Gehant, S.; Gerritsen, V.; Gos, A.; Gruaz-Gumowski, N.; Hinz, U.; Hulo, C.; James, J.; Jungo, F.; Keller, G.; Lara, V.; Lemercier, P.; Lew, J.; Lieberherr, D.; Lombardot, T.; Martin, X.; Masson, P.; Morgat, A.; Neto, T.; Paesano, S. *Nucleic acids research* **2014**, *42*, D191-D198.

(191); 2013.08 ed.; Chemical Computing Group Inc.: Montreal, QC, Canada, 2013.

(192)Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. *Journal of molecular graphics & modelling* **2006**, *25*, 247-260.

(193)Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J Chem Phys* **1983**, *79*, 926-935.

(194)Hanson, M. A.; Cherezov, V.; Griffith, M. T.; Roth, C. B.; Jaakola, V. P.; Chien, E. Y.; Velasquez, J.; Kuhn, P.; Stevens, R. C. *Structure* **2008**, *16*, 897-905.

(195)Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J Comput Chem* **2004**, *25*, 1157-1174.

(196)Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J Comput Phys* **1977**, *23*, 327-341.

(197)Miyamoto, S.; Kollman, P. A. *J Comput Chem* **1992**, *13*, 952-962.

(198)Roe, D. R.; Cheatham, T. E. *J Chem Theory Comput* **2013**, *9*, 3084-3095.

(199)D.A. Case, T. A. D., T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A.W. Götz, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, R.M. Wolf, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P.A. Kollman University of California, San Francisco, 2012.

(200)Humphrey, W.; Dalke, A.; Schulten, K. *Journal of molecular graphics* **1996**, *14*, 33-38, 27-38.

(201)Stone, J., University of Missouri, Rolla, 1998.

(202)Frishman, D.; Argos, P. *Proteins* **1995**, *23*, 566-579.

(203)Varshney, A. B., F. P.; Wright W. V. *IEEE Computer Graphics and Applications* **1994**, *14*, 19-25.

(204)Kim, H. J.; Lim, H. H.; Rho, S. H.; Eom, S. H.; Park, C. S. *The Journal of biological chemistry* **2006**, *281*, 38573-38581.

(205)Tanford, C. *J Am Chem Soc* **1962**, *84*, 4240-4247.

(206)White, S. H.; Wimley, W. C. *Biochimica et biophysica acta* **1998**, *1376*, 339-352.

(207)Lu, R.; Alioua, A.; Kumar, Y.; Eghbali, M.; Stefani, E.; Toro, L. *The Journal of physiology* **2006**, *570*, 65-72.

(208)Stefani, E.; Ottolia, M.; Noceti, F.; Olcese, R.; Wallner, M.; Latorre, R.; Toro, L. *Proceedings of the National Academy of Sciences of the United States of America* **1997**, *94*, 5427-5431.

(209)Jaggar, J. H.; Porter, V. A.; Lederer, W. J.; Nelson, M. T. *American journal of physiology. Cell physiology* **2000**, *278*, C235-256.

(210)Chang, T.; Wu, L.; Wang, R. *American journal of hypertension* **2006**, *19*, 678-685.

(211)Petkov, G. V.; Bonev, A. D.; Heppner, T. J.; Brenner, R.; Aldrich, R. W.; Nelson, M. T. *The Journal of physiology* **2001**, *537*, 443-452.

(212)Ponte, C. G.; McManus, O. B.; Schmalhofer, W. A.; Shen, D. M.; Dai, G.; Stevenson, A.; Sur, S.; Shah, T.; Kiss, L.; Shu, M.; Doherty, J. B.; Nargund, R.; Kaczorowski, G. J.; Suarez-Kurtz, G.; Garcia, M. L. *Mol Pharmacol* **2012**, *81*, 567-577.

(213)Bolton, E. E.; Wang, Y.; Thiessen, P. A.; Bryant, S. H. In *Annual Reports in Computational Chemistry*; Ralph, A. W., David, C. S., Eds.; Elsevier: 2008; Vol. Volume 4, p 217-241.

(214)Halgren, T. A. *J Comput Chem* **1996**, *17*, 490-519.

(215)Bukiya, A. N.; Patil, S. A.; Li, W.; Miller, D. D.; Dopico, A. M. *ChemMedChem* **2012**, *7*, 1784-1792.

(216)Patil, S.; Bukiya, A. N.; Li, W.; Dopico, A. M.; Miller, D. *Bioorg Med Chem Lett* **2008**, *18*, 3427-3430.

(217)Labute, P. *J Chem Inf Model* **2010**, *50*, 792-800.

(218)Triballeau, N. B., H.O.; Acher, F. *Pharmacophores and Pharmacophore Searches*; Wiley-VCH: Weinheim, Germany, 2006; Vol. 32.

(219); MDL Information Systems, Inc.: San Leandro, CA.

(220)Rogers, D. J.; Tanimoto, T. T. *Science* **1960**, *132*, 1115-1118.

(221)Chen, Y.; Salem, R. M.; Rao, F.; Fung, M. M.; Bhatnagar, V.; Pandey, B.; Mahata, M.; Waalen, J.; Nievergelt, C. M.; Lipkowitz, M. S.; Hamilton, B. A.; Mahata, S. K.; O'Connor, D. T. *American journal of nephrology* **2010**, *32*, 414-424.

(222)Kang, C.; Li, Q. *Curr Opin Chem Biol* **2011**, *15*, 560-569.

(223)Sastry, M.; Bewley, C. A.; Kwong, P. D. *Advances in experimental medicine and biology* **2012**, *992*, 197-211.

(224)Hannig, G.; Makrides, S. C. *Trends in biotechnology* **1998**, *16*, 54-60.

(225)Stols, L.; Zhou, M.; Eschenfeldt, W. H.; Millard, C. S.; Abdullah, J.; Collart, F. R.; Kim, Y.; Donnelly, M. I. *Protein expression and purification* **2007**, *53*, 396-403.

(226)Drew, D.; Froderberg, L.; Baars, L.; de Gier, J. W. *Biochimica et biophysica acta* **2003**, *1610*, 3-10.

(227)Page, R. C.; Moore, J. D.; Nguyen, H. B.; Sharma, M.; Chase, R.; Gao, F. P.; Mobley, C. K.; Sanders, C. R.; Ma, L.; Sonnichsen, F. D.; Lee, S.; Howell, S. C.; Opella, S. J.; Cross, T. A. *Journal of structural and functional genomics* **2006**, *7*, 51-64.

(228)Yang, Z.; Zhang, L.; Zhang, Y.; Zhang, T.; Feng, Y.; Lu, X.; Lan, W.; Wang, J.; Wu, H.; Cao, C.; Wang, X. *Plos One* **2011**, *6*, e22981.

(229)Sheibani, N. *Prep Biochem Biotechnol* **1999**, *29*, 77-90.

(230)Stols, L.; Gu, M.; Dieckman, L.; Raffen, R.; Collart, F. R.; Donnelly, M. I. *Protein expression and purification* **2002**, *25*, 8-15.

(231)Haun, R. S.; Serventi, I. M.; Moss, J. *BioTechniques* **1992**, *13*, 515-518.

(232)Zuo, X.; Li, S.; Hall, J.; Mattern, M. R.; Tran, H.; Shoo, J.; Tan, R.; Weiss, S. R.; Butt, T. R. *Journal of structural and functional genomics* **2005**, *6*, 103-111.

(233)Miroux, B.; Walker, J. E. *Journal of molecular biology* **1996**, *260*, 289-298.

(234)Dumon-Seignovert, L.; Cariot, G.; Vuillard, L. *Protein expression and purification* **2004**, *37*, 203-206.

(235)Marassi, F. M.; Opella, S. J. *Current opinion in structural biology* **1998**, *8*, 640-648.

(236)Borch, J.; Hamann, T. *Biological chemistry* **2009**, *390*, 805-814.

(237)Mus-Veteau, I. *Heterologous expression of membrane proteins : methods and protocols*; Humana Press: New York, 2010.

(238)Maniatis, T.; Fritsch, E. F.; Sambrook, J. *Molecular cloning : a laboratory manual*; Cold Spring Harbor Laboratory: Cold Spring Harbor, N.Y., 1982.

(239)Benson, D. A.; Clark, K.; Karsch-Mizrachi, I.; Lipman, D. J.; Ostell, J.; Sayers, E. W. *Nucleic acids research* **2014**, *42*, D32-37.

(240)Artimo, P.; Jonnalagedda, M.; Arnold, K.; Baratin, D.; Csardi, G.; de Castro, E.; Duvaud, S.; Flegel, V.; Fortier, A.; Gasteiger, E.; Grosdidier, A.; Hernandez, C.; Ioannidis, V.; Kuznetsov, D.; Liechti, R.; Moretti, S.; Mostaguir, K.; Redaschi, N.; Rossier, G.; Xenarios, I.; Stockinger, H. *Nucleic acids research* **2012**, *40*, W597-603.

(241)Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. *Journal of molecular biology* **1990**, *215*, 403-410.

(242)Mullis, K.; Faloona, F.; Scharf, S.; Saiki, R.; Horn, G.; Erlich, H. *Cold Spring Harbor symposia on quantitative biology* **1986**, *51 Pt 1*, 263-273.

(243)Saiki, R. K.; Gelfand, D. H.; Stoffel, S.; Scharf, S. J.; Higuchi, R.; Horn, G. T.; Mullis, K. B.; Erlich, H. A. *Science* **1988**, *239*, 487-491.

(244)Laemmli, U. K. *Nature* **1970**, *227*, 680-685.

(245)Gaetjens, E.; Pertschuk, L. P. *Journal of steroid biochemistry* **1980**, *13*, 1001-1003.

(246)Rath, A.; Glibowicka, M.; Nadeau, V. G.; Chen, G.; Deber, C. M. *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106*, 1760-1765.

(247)Huang, C.; Mohanty, S.; Banerjee, M. *Biochemistry* **2010**, *49*, 1115-1126.

(248)Tamm, L. K.; Liang, B. Y. *Progress in nuclear magnetic resonance spectroscopy* **2006**, *48*, 201-210.

(249)Wroblewski, H.; Burlot, R.; Johansson, K. E. *Biochimie* **1978**, *60*, 389-398.

(250)Curtis-Fisk, J.; Spencer, R. M.; Weliky, D. P. *Protein expression and purification* **2008**, *61*, 212-219.

(251)Ohe, M.; Kajita, A. *Biochemistry* **1980**, *19*, 4443-4450.

(252)Casal, H. L.; Kohler, U.; Mantsch, H. H. *Biochimica et biophysica acta* **1988**, *957*, 11-20.

(253)Robichon, C.; Luo, J. Y.; Causey, T. B.; Benner, J. S.; Samuelson, J. C. *Appl Environ Microb* **2011**, *77*, 4634-4646.

(254)Stanczak, P.; Horst, R.; Serrano, P.; Wuthrich, K. *J Am Chem Soc* **2009**, *131*, 18450-18456.

(255)Roda, A.; Minutello, A.; Angellotti, M. A.; Fini, A. *J Lipid Res* **1990**, *31*, 1433-1443.

(256)Johnsson, B.; Lofas, S.; Lindquist, G. *Anal Biochem* **1991**, *198*, 268-277.

(257)Narasinga Rao, M. S. *J Am Chem Soc* **1962**, *84*, 1788-1790.

(258)Karaboga, A. S.; Petronin, F.; Marchetti, G.; Souchet, M.; Maigret, B. *Journal of molecular graphics & modelling* **2013**, *41*, 20-30.

(259)Warschawski, D. E.; Arnold, A. A.; Beaugrand, M.; Gravel, A.; Chartrand, E.; Marcotte, I. *Biochimica et biophysica acta* **2011**, *1808*, 1957-1974.

Appendix A: DNA and Protein Sequence of *E. Coli* Optimized β_1

```
CAT ATG ATG GGC AAA AAA CTG GTT ATG GCA CAG AAA CGT GGT GAA
 H   M   M   G   K   K   L   V   M   A   Q   K   R   G   E

ACC CGT GCA CTG TGT CTG GGT GTT GCA ATG GTT GTT TGT GCA GCA
 T   R   A   L   C   L   G   V   A   M   V   V   C   A   A

ATT ACC TAT TAT ATC CTG GGC ACC ACC GTT CTG CCG CTG TAT CAG
 I   T   Y   Y   I   L   G   T   T   V   L   P   L   Q   K

AAA AGC GTT TGG ACC CAA GAA AGC ACC TGT CTG CTG GTT GAA ACC
 K   S   V   W   T   Q   E   S   T   C   L   L   V   E   T

AAC ATT AAA GAT CAA GAA GAA CTG GAA GGT CGT AAA GTT CCG CAG
 N   I   K   D   Q   E   E   L   E   G   R   K   V   P   Q

TAT CCG TGT CTG TGG GTT AAT GTT AGC GCA GTT GGT CGT TGG GCA
 Y   P   C   L   W   V   N   V   S   A   V   G   R   W   A

ATG CTG TAT CAT ACC GAA GAT ACC CGT GAT CAG AAT CAG CAG TGT
 M   L   Y   H   T   E   D   T   R   D   Q   N   Q   Q   C

AGC TAT ATT CCG CGT AAT CTG GAT AAC TAT CAG ACC GCA CTG GTT
 S   Y   I   P   R   N   L   D   N   Y   Q   T   A   L   V

GAT GTT AAA AAA GTT CGT GCC AAC TTC TAC AAA CAC CAC AAC TTT
 D   V   K   K   V   R   A   N   F   Y   K   H   H   N   F

TAT TGT TTT AGC GCA CCG CAG GTT AAT GAA ACC AGC GTT GTT TAT
 T   C   F   S   A   P   Q   V   N   E   T   S   V   V   Y

CAG CGT CTG TAT GGT CCG CAG ATT CTG CTG TTT AGC TTT TTT TGG
 Q   R   L   Y   G   P   Q   I   L   L   F   S   F   F   W

CCG ACC TTT CTG CTG ACC GGT GGT CTG CTG ATT ATT GCC ATG GTT
 P   T   F   L   L   T   G   G   L   L   I   I   A   M   V

AAA CTG AAT CGT AGC CTG AGC GTT CTG GCA GCA CAG AAA TAA CTC
 K   L   N   R   S   L   S   V   L   A   A   Q   K Stop  L

GAG
 E
```